GEOG 281: The World & Big Data

Fall 2020

Instructor	Dr. Hui (Henry) Luan hluan@uoregon.edu Office hours: Mondays, 10am-12pm
Lecture	Tuesdays & Thursdays, 8am–9am
Labs	Section 1: Thursdays, 10am-11am Section 2: Fridays, 10am-11am
GE	Michala Garrison michalag@uoregon.edu Office hours: Mondays, 1pm-3pm Eric Levenson elevens2@uoregon.edu Office hours: Wednesdays, 10am-12pm

Course Description

Today's world is driven by data. Political, economical, and environmental decision-making is increasingly dependent on big data sets. Services like Google personalize what you see and advertise upon data about you and 'customers like you'; social media allow the public to directly and proactively participate in activities such as urban planning, disaster reporting, and humanitarian interventions.

"80% of data is geographic" (Morais, 2012). While this statement is arguable, real-world datasets are geographically and temporally referenced when they are collected by devices with location-aware sensors such as the Global Positioning System (GPS) receivers, mobile phones, and vehicles. Technologies such as wireless local area networks (WLAN) make it feasible to collect georeferenced data within the indoor environments. The increasing volume and varying formats of these geospatial big spatial datasets have posed new challenges in data storage, management, process, analysis, visualization, and quality control. Ethical and geo-privacy issues have also emerged in recent applications of big data techniques in tracking human activities.

This course will introduce relevant concepts (e.g., 5 V's) and techniques (e.g., cloud computing) of big (spatial) data as well as its applications in the real word, such as building smart cities as well as disaster management. The students will also have hands-on experiences such as contributing Volunteered Geographic Information (VGI, e.g., creating and uploading geographic information to Open Street Map), retrieving social media data associated with location (e.g., "geo-tagged" Tweets), operating relational database management system (RDMS) and NoSQL databases, and visualizing big spatial datasets.

Textbooks

1. V. Mayer-Schönberger and K. Cukier. *Big data: a revolution that will transform how we live, work, and think*. Eamon Dolan/Mariner Books, 2014.
2. Foster I, Ghani R, Jarmin RS, Kreuter F, and Lane J (Eds.). *Big data and social science: A practical guide to methods and tools*. CRC Press, 2017. (Freely accessible from UO library)
3. (Optional) Thatcher J, Eckert J, and Shears A (Eds.). *Thinking big data in geography: new regimes, new research*. University of Nebraska Press, 2018. (Freely accessible from UO library)

Learning objectives

After completing this course, the students are expected to:

- Understand the unique characteristics of big data
- Be familiar with the techniques used for collecting, storing, managing, processing, analyzing, and visualizing big data
- Know how big data can solve real-world problems
- Grasp basic skills to contribute, retrieve, and visualize big spatial datasets
- Be aware of issues in using big spatial data

Course outline

Week 1		
	Sep 29	**Introduction to the course**
	Oct 1	**What is big (spatial) data?**
	Reading	1. Mayer-Schönberger and Cukier (2014): Chapter 1: NOW & Chapter 5: DATAFICATION 2. Foster et al. (2017): Chapter 1: Introduction
Week 2		
	Oct 6	**The 5 V's of big data**
	Oct 8	**Spatial data types**
	Reading	1. Mayer-Schönberger and Cukier (2014): Chapter 2: MORE
Week 3		
	Oct 13	**Sources of big spatial data**
	Oct 15	**Sources of big spatial data (cont'd)**
	Reading	1. Foster et al. (2017): Chapter 2: Working with Web Data and APIs, pp. 23-27 2. Chi M. et al. Big data for Remote Sensing: Challenges and Opportunities, *Proceedings of the IEEE*, 2016, 104(11).

Week 4		
	Oct 20	**Retrieving social media data: R basics**
	Oct 22	**Retrieving social media data: Twitter developer account**
	Reading	1. Lingel J. (2017) Social Media. In Kitchin R, Lauriault TP, and Wilson MW (Eds.), Understanding Spatial Media (pp.66-73). London, SAGE Publications Ltd. 2. Foster et al. (2017): Chapter 7: Text analysis
Week 5		
	Oct 27	**Big data quality I**
	Oct 29	**Big data quality II**
	Reading	1. Mayer-Schönberger and Cukier (2014): Chapter 3: MESSY 2. Foster et al. (2017): Chapter 2: Working with Web Data and APIs – Quality, scope, and management, pp. 44-46 3. Foster et al. (2017): Chapter 10: Errors and Inference
Week 6		
	Nov 3	**Mid-term exam**
	Nov 5	**Basics in (spatial) database**
	Reading	1. Foster et al. (2017): Chapter 4: Databases
Week 7		
	Nov 10	**Big spatial data management and storage**
	Nov 12	**Visualizing big data**
	Reading	1. Foster et al. (2017): Chapter 9: Information Visualization
Week 8		
	Nov 17	**Artificial intelligence for handling big data I**
	Nov 19	**Artificial intelligence for handling big data II**
	Reading	1. Mayer-Schönberger and Cukier (2014):Chapter 6: VALUE & Chapter 7: IMPLICATIONS 2. Foster et al. (2017): Chapter 5: Programming with Big Data 3. (Optional) Foster et al. (2017): Chapter 6: Machine Learning
Week 9		
	Nov 24	**Geo-privacy and ethical issues**
	Nov 26	**No class: Thanksgiving vacation**

Reading		1. Mayer-Schönberger and Cukier (2014): Chapter 8: RISKS & Chapter 9: CONTROL 2. Foster et al. (2017): Chapter 11: Privacy and Confidentiality
Week 10		
Dec 1	**Group project presentation**	
Dec 3	**Group project presentation (cont'd)**	

Assignments

Assignments	**Available date on Canvas**	**Due date (midnight)**
1: Contributing VGI using OpenStreetMap & Examining landscape change using Google Earth	**Oct 8**	**Oct 23**
2: Retrieving and analyzing (geo-tagged) Tweets using R	**Oct 22**	**Nov 6**
3: RDBM (Microsoft Access) & NoSQL (MongoDB)	**Nov 5**	**Nov 20**
4: Geo-visualizing big accident datasets in the UK using Kepler.gl	**Nov 19**	**Nov 27**

Group project

Requirements	**Submission/Output**
▪ 7-8 students per group (group on your own or randomly assigned by the GE) ▪ Identify an existing big spatial data project (domestic or international) ▪ Talk to GE before presentation to avoid duplicates **(due Nov 6)**	▪ 15 mins presentation ▪ 5-page report (Times New Roman, 12pt, double-space) **(due Dec 7, midnight)**

Grading

Assignments	**40% (10%*4)**
Group project	**25% (Report 15% + presentation 10%)**
Midterm exam	**15%**
Final exam	**20%**

Notes:

- Late assignments will receive a deduction of 5% per day including weekends and holidays. Assignments submitted 7 days later than the deadline will NOT be accepted. The only exception is when you can provide a university-approved excuse.
- All times are based on Pacific Standard Time (PST).
- Grades for the course will be based on the following grading scale.

A+ 97-100	B+ 85-89.9	C+ 70-74.9	D 55-59.9
A 92-96.9	B 80-84.9	C 65-69.9	F less than 55
A- 90-91.9	B- 75-79.9	C- 60-64.9	

Academic Integrity Code

All students are expected to complete assignments in a manner consistent with academic integrity. Students must produce their own work and properly acknowledge and document all sources (ideas, quotations, paraphrases). Students can find more complete information about the University of Oregon's Policy on Academic Dishonesty in the University of Oregon Student Handbook. **If you are found to have plagiarized (copied) off a classmate or from other materials for a test or an assignment of any sort, your first warning will be a zero on the assignment. Your second incident will result in an F in the course.**

Accessibility Statement

The University of Oregon is working to create inclusive learning environments. Please notify me at the beginning of the term if there are aspects of the instruction or design of this course that result in disability-related barriers to your participation. You are also encouraged to contact the Accessible Education Center in 164 Oregon Hall at 541-346-1155 or uoaec@uoregon.edu.

The professor of this course reserves the right to change aspects of this syllabus any time during the term. Students will be informed if and when this occurs.