COURSE OUTLINE
Data Visualization and Analytics in East Asian Studies

The ubiquitous encounter with digitized and born-digital sources has become a quotidian experience for the present-day researcher of East Asian studies. What are the implications of the ongoing "digital turn" in the humanities? In what ways does our age of information abundance influence the craft, modes, and expectations of research in our trade?

This lab course introduces students to visual and exploratory analysis of humanities data germane to East Asia, both preindustrial and modern, offering technical training and critical reflections in three modules. The initial sessions begin with discussions about the burgeoning meta-discipline known as "the digital humanities" and we will examine the meteoric rise and pervasive presence of "the digital" in East Asian societies. We then will proceed with lab sessions on the management, visualization, and analytics of various types of humanities data. In the final module, the focus will shift to methodological ruminations about the virtues and shortcomings of computational methods.

Module A. Introduction to "Digital Humanities"
"the digital turn": pedagogy, research, publication, social media, open access
from humanities computing to digital humanities: a history going back to the 1960s
transition to "digital humanities 2.0": metaLAB at Harvard, beyond "data buckets"
East Asia goes digital: "digital" as national identity, digitization of sources, techno-utopianism

Module B: Building and Visualizing Databases
databases: structured and unstructured data, variables, data types, data and metadata
relational database: data tables, relationships, SQL query
XML: markup languages, hierarchical structure, XSLT, XPath, XQuery
spatial data: vector and raster, points, lines, polygons, basics of GIS
biographical data: from unstructured pool to relational databases and tagged documents
network data: graph theory, aggregate vs. centrality
network visualization: edge list, node attribute table,
network type: random vs. small world vs. organic vs. complex
network topology: line, ring, mesh
data mining: regular expressions, GATE
topic modeling: MALLET

Module C: Critical Reflections
distant reading: Franco Moretti
speculative computing: Johanna Drucker
sampling: Lev Manovich, how to manage "big data"
scientific visualization vs. humanities visualization
programming the humanities: should humanists code? how to handle born-digital sources
data mining and topic modeling: caveats and limitations of automated information excavation
authoring studies in new media
lessons from archaeology: from processualism to post-processualism to neo-pragmatism