

## CSCI 680/490 Visual Analytics - Midterm Exam

### Instructions

This exam is a take-home exam with 2 required questions. You are not allowed to discuss the exam with anyone, particularly other students in the class. If you have any questions, please contact the professor (Maoyuan Sun, [smaoyuan@niu.edu](mailto:smaoyuan@niu.edu)). Any problems/questions with the exam should be reported to me directly.

Submit a single ZIP file with a MS Word document (including your written answers to question 1 and 2), and code files (for questions 2). Use file names or a directory structure to clearly indicate what is what.

The exam must be submitted in Blackboard by **11:59pm on March 07, 2021**.  
**NO LATE** exam submissions will be accepted.

### Q1: Essay (50 points)

Read “Defining Insight for Visual Analytics” by Chang et al.

1. Summarize what are insights of visual analytics defined in this paper?
2. This article was published in 2009. As the rapid rising of data mining and machine learning techniques, are there any important changes on the insights of visual analytics? Make an argument for this. If there are, describe the changes. If there are not, discuss your reasons. Use examples from class or elsewhere to support your argument.

You can find the paper in the attachment.

### Q2: Development & Analysis (50 points)

Given the car dataset in the attachment, you need to do a visual analysis of clustering these cars.

1. You should pick **at least** four columns for clustering the cars. You are free to decide which columns to use and determine your rules for the clustering.
2. Using D3.js to visualize your clustering results.
  - a. Use a circle to represent each car
  - b. Cars in the same cluster should be near each other
  - c. Use different colors for different clusters
  - d. When mouse-hovering or selecting a car, you should show its detailed information from the given dataset.
3. You should answer the following questions, besides submitting your code.
  - a. Which columns do you choose for performing the clustering?
  - b. What is your rule for deciding whether a car belongs to a cluster or not?
  - c. How many cars are there in the largest cluster?
  - d. Do you find any car(s) falsely clustered? For example, it is more reasonable that *car A* should be in *cluster 2*, but it is displayed in *cluster 1*. If so, you need to explain why this happens.

Regarding to the development part, you should use good software engineering practices. Comment your code, use consistent formatting, use meaningful variable names, etc.