Introduction to Big Data

Team Project Guideline

Keungoui Kim
awekim@handong.edu

- Choose a "specific" socio-economic topic and key issue that all the team members are interested in.
  - (general) inequality, poverty, aging, COVID, election, real-estate, etc.
- Collect data relevant to the topic
- **Conduct data analysis to explain selected issue with data-oriented perspective**

- The goal of this team project is to practice using problem-based methodologies to find data that match the selected topic and to get key messages and interpretations from the data

- **Introduction** (15pts)
  - Background (5 pts) → Is the motivation of this project sufficiently explained?
  - Problem statement (5 pts) → Is the project problem specific and clear enough?
  - Purpose & goal (5 pts) → Are the purpose and goal clearly stated?
- **Data preparation** (30 pts)
  - Data searching (10 pts) → Are they found the appropriate data set for the project?
  - Data description (10 pts) → Is the data description sufficient to understand the data?
  - Data preprocessing (10 pts) → Are they conducted all necessary data preprocessing task?
- **Data analysis** (30 pts)
  - Analytic method (10 pts) → Do they select a proper method? Do they fully understand data analysis method?
  - Visualization (10 pts) → Are the visualized works easy to understand? Do they help you understand the project?
  - Interpretation (10 pts) → Do they provide correct interpretation?
- **Conclusion** (10 pts)
  - Concluding remark → Do you think the project give you the new intuition?

- **Final presentation** Week 15
  - Introduction: background + problem statement + purpose + goal
  - Data preparation: data searching + description + preprocessing
  - Data analysis: analytic method + visualization + interpretation
  - Conclusion: concluding remark
  - Reference
  - Time duration: 10 minutes

- Submission
  - All files used for the project including presentation, .ipynb file, data, etc.
  - Submit to HDLMS