Purdue University Northwest
Department of Electrical and Computer Engineering
Introduction to Visualization Techniques
Fall 2018

Project: Multidimensional Data Visualization

- **Dataset**

I encourage you to explore any dataset you find interesting unless it is inappropriate or too small/ too simple for this class. **It is helpful to write down a brief description of 1) your data file, 2) your design and 3) the visualization packages/libraries if you use a third-party software package before you start programming.**

- The description of data files should contain the name and a brief explanation of the dataset (what it is), the source (where you get it), the fields (the dimensional information) and the number of records (the size). You can convert the format or revise the structure of the dataset if necessary.
- The description of visualization toolkits/packages/libraries should contain justifications (why it is a good choice for your data), and the programming language. Check with me if you have any questions.
- The description of your design should contain a representative sketch and a brief description.

I list several data repositories below for your reference.

1) Internet traffic data: http://ita.ee.lbl.gov/html/traces.html
2) Visual Analytics Benchmark Repository: http://hcil2.cs.umd.edu/newvarepository/
3) Google Public Data Directory: https://www.google.com/publicdata/directory
4) Stanford Large Network Dataset Collection: http://snap.stanford.edu/data/index.html
5) Amazon AWS Public Datasets: https://aws.amazon.com/public-datasets/
6) A good article on Quora including links to a variety of data repositories: https://www.quora.com/Data/Where-can-I-find-large-datasets-open-to-the-public
7) Headoop Illuminated book: http://hadoopilluminated.com/hadoop_illuminated/Public_Bigdata_Sets.html
   Dataset links are grouped in categories: Generic Repositories, Geo data, Web data, and Government data.

- **Report**

You should briefly describe your findings in the data file and link the findings to your design and implementation. The project can be implemented in any visualization techniques and toolkits. It is recommended to use one visualization tool/toolkit only. Your project and report should make interesting findings, show new visualization effects, develop user-friendly features and compare with other multidimensional visualization techniques (e.g. parallel coordinates). You may include a short section describing the experience of the toolkit you use. For example, the problems that you have encountered, the functions that you feel convenient to use and the features you want.

Submit the report, the presentation and your code in a zip file (not in .rar, .7z etc).

- **Presentation**

Each group should prepare a 15 minute presentation (including 2 minutes for questions) with following sections:

1) Introduction: The dataset, your findings, the visualization toolkit or library (2-3 slides)

2) Implementation: deign, visualization methods and effects, toolkit functions and development details. (3-6 slides)

3) Demonstration: representative screenshots and comparisons (3-6 slides) and demonstration.

4) Discussions and conclusions: summaries, toolkit comments. (2-4 slides)