San José State University

Math 253: Mathematical Methods for Data Visualization

Lecture 3: High quality data plotting in MATLAB

Dr. Guangliang Chen

Outline

- **Focus of this lecture**: Data plotting and exploration in MATLAB
  - Data types
  - Storing data
  - Importing data
  - Plotting data
  - Exploring data
- **Learning resources**:

  - 2-D and 3-D Plots[1]
  - Types of MATLAB Plots [2]
  - MATLAB Plot Gallery[3]

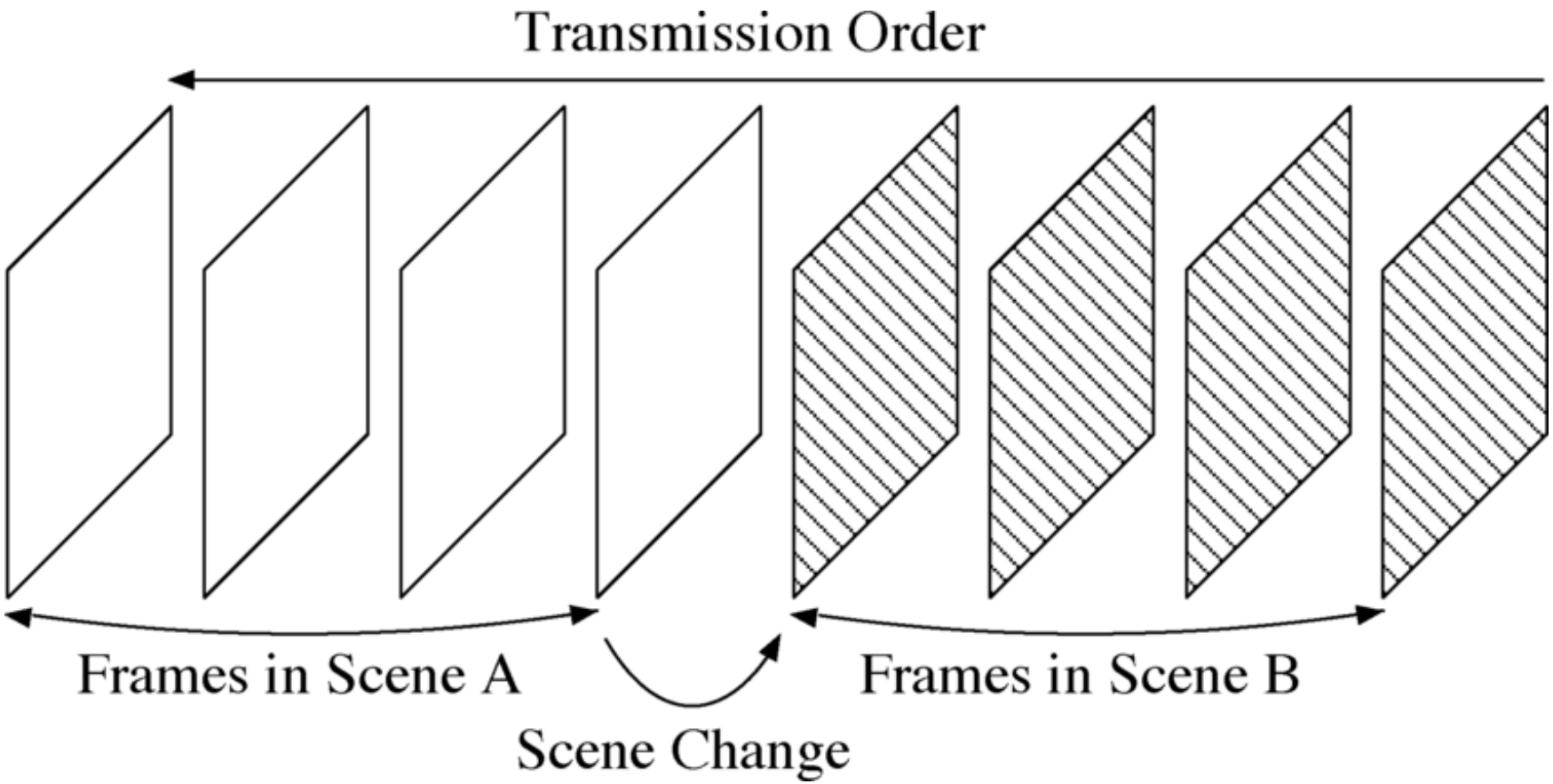
- **HW3** (plotting): due Thursday, 2/20, in class

[1] https://www.mathworks.com/help/matlab/2-and-3d-plots.html
[2] https://www.mathworks.com/help/matlab/creating_plots/types-of-matlab-plots.html
[3] https://www.mathworks.com/products/matlab/plot-gallery.html

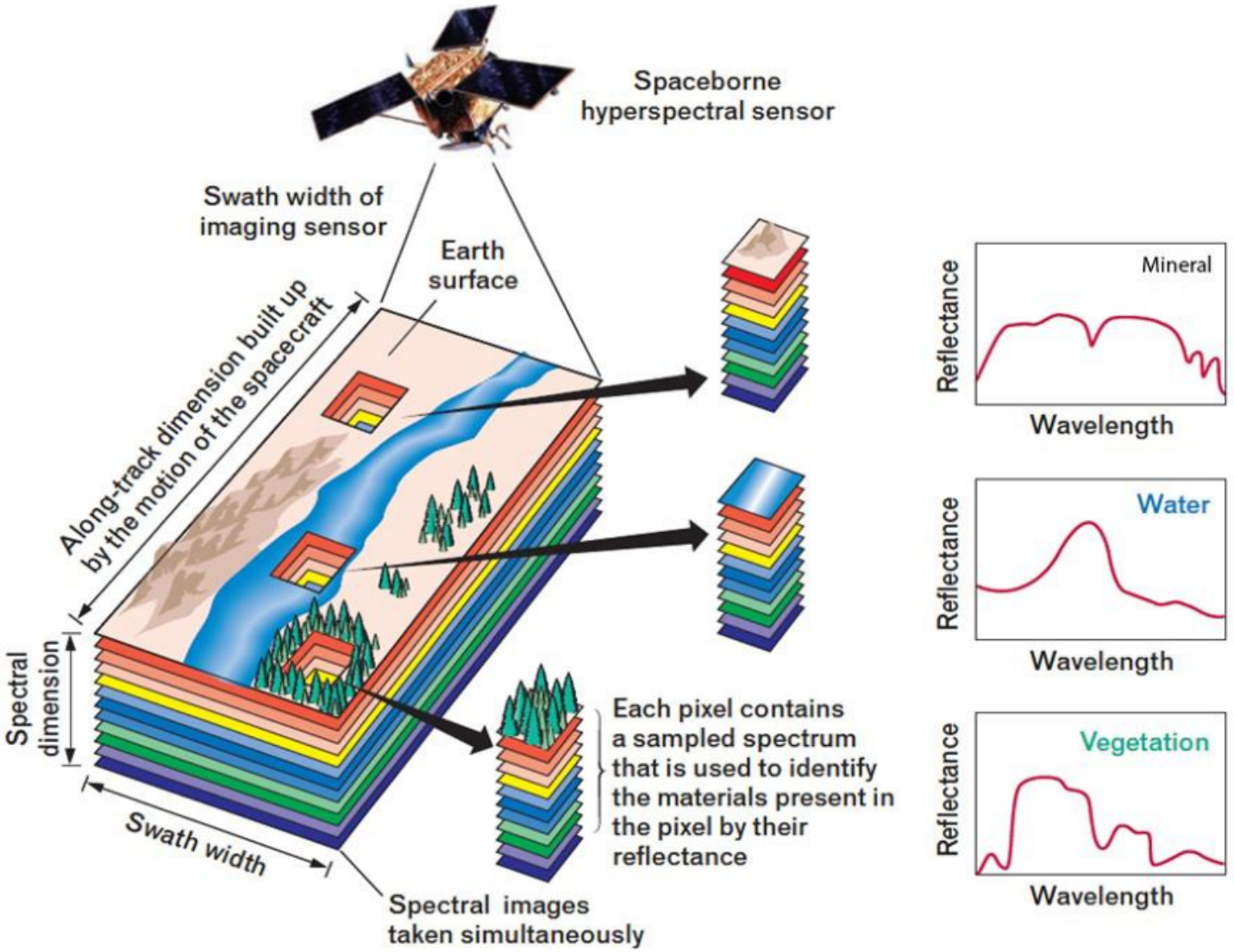
Data types

Data exists (or is collected) in different forms:

- Numerical (vectors)
- Categorical
- Graphs (networks)
- Text documents
- Images (gray-scale, color)
- **Videos**
- **Hyperspectral images**


Transmission Order
Frames in Scene A
Frames in Scene B
Scene Change

High quality data plotting in MATLAB



Storing data as arrays

In MATLAB, data sets are typically stored as arrays:

- **1-D array (vector)**: measurements of a single quantitative variable, e.g., height $= (170, 183, 178, 165, 175)$
- **1-D cell array**: measurements of a single categorical variable, e.g., sex $= \{`F', `M', `M', `F', `F'\}$
- **2-D array (matrix)**: measurements of multiple quantitative variables, digital images (single or collection), text corpus, transition probabilities of a Markov chain
- **3-D array**: collections of images, video sequences, hyperspectral images

Main data sets for demonstration

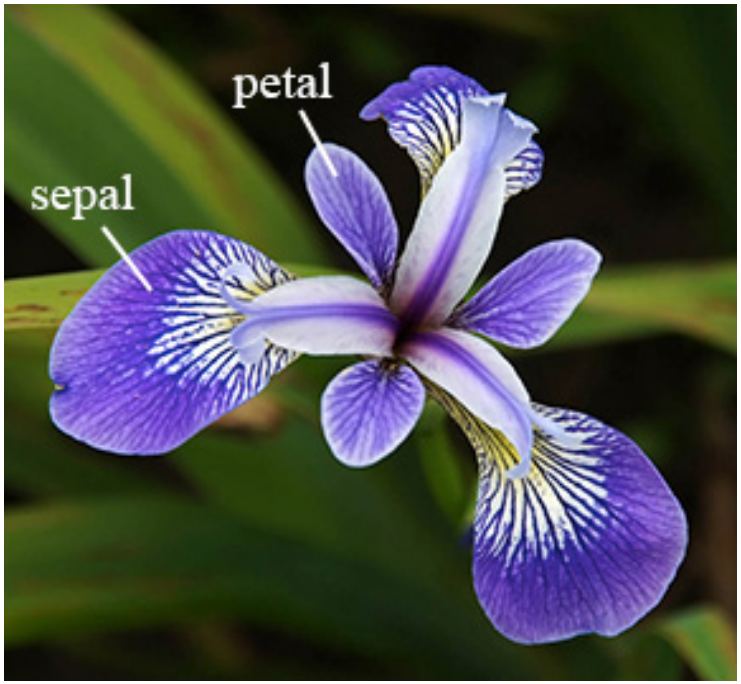
- UCI Machine Learning Repository
  - Iris data[4]
  - Wine quality[5]
- MNIST handwritten digits[6]
- 20 Newsgroups[7]

[4] https://archive.ics.uci.edu/ml/datasets/Iris
[5] https://archive.ics.uci.edu/ml/datasets/wine+quality
[6] http://yann.lecun.com/exdb/mnist/
[7] http://qwone.com/~jason/20Newsgroups/

The Iris Data Set (created by R.A. Fisher)

Dataset information:

- **150 instances**
- **4 numerical attributes**
  - sepal length in cm
  - sepal width in cm
  - petal length in cm
  - petal width in cm
- **1 categorical variable: class** (Iris Setosa, Iris Versicolour, Iris Virginica)



The Wine Quality Data Set

Dataset information:

- **4,898 instances** (two datasets are included, related to red and white vinho verde wine samples, from the north of Portugal)
- **11 numerical attributes** (based on physiochemical tests)
- **1 output variable: quality** (score between 0 and 10)

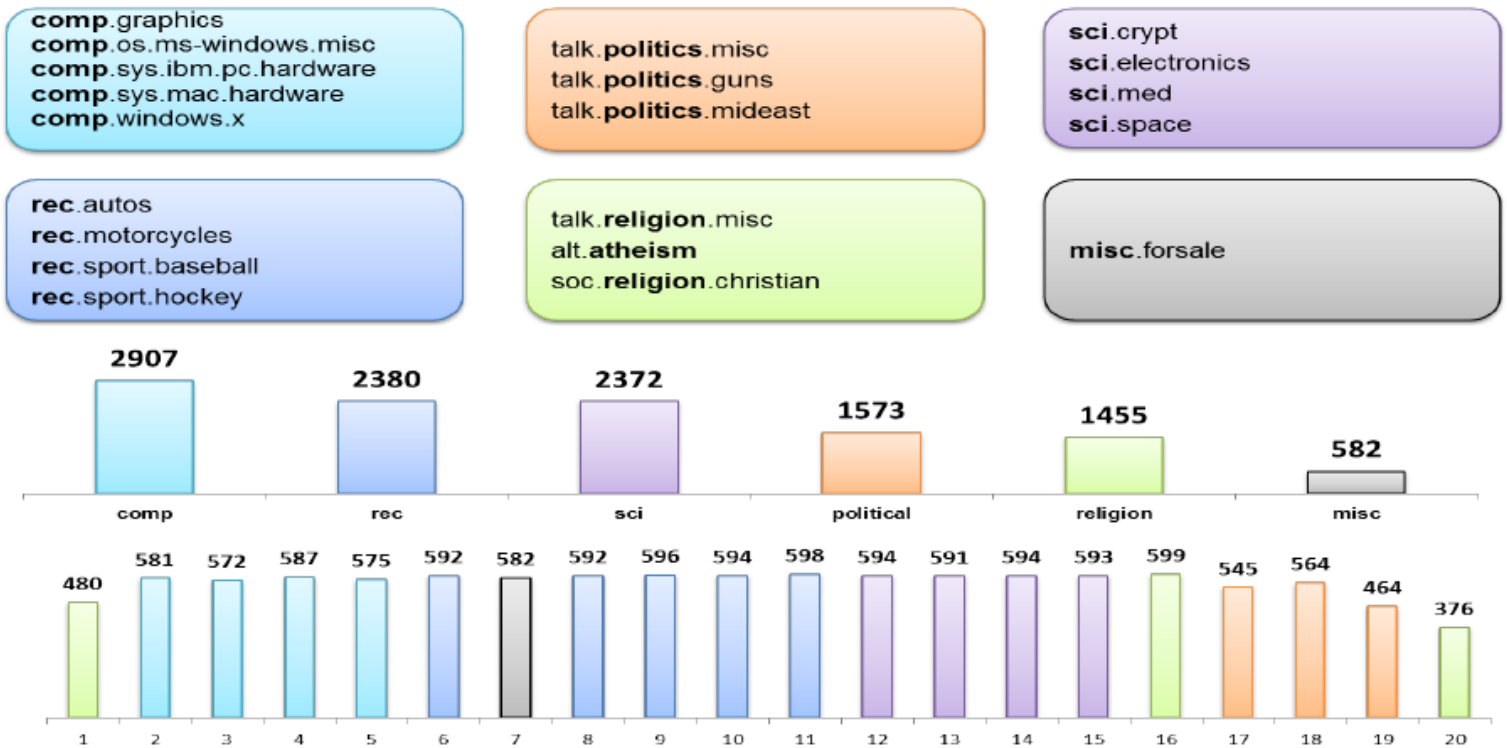
MNIST Handwritten Digits

It is a benchmark data set for machine learning (by Yann LeCun of Facebook), consisting of 70,000 handwriting examples of approximately 250 writers:

- Black/white images of size 28×28 each
- 60,000 for training and 10,000 for testing

20 Newsgroups Data Set

A collection of nearly 20,000 newsgroup documents, partitioned (approximately) evenly across 20 different newsgroups:



Data visualization

Goals: For each data set, we will focus on both of the following

- data exploration (for insights)
- data plotting (with publication quality)

Strategy: We will examine the variables in the following ways:

- Single variable:
  - Numerical: 1-D scatterplot, histogram, boxplot, bar graph (if frequency data)
  - Categorical: bar graph, pie chart

High quality data plotting in MATLAB

- Two variables:
  - Both numerical: 2-D scatterplot
  - Both categorical: stacked bar plot
  - Mixed: side-by-side boxplot
- Three variables:
  - All numerical: 3-D scatterplot, scatterplot matrix
  - Two numerical and 1 categorical: 2-D scatterplot with groups
  - One numerical and two categorical: heatmap, 3D bar plot

In-class demonstrations

See scripts from instructor in class.