Data Visualization and Computational Art History

Lev Manovich / Software Studies Initiative
Visual Art Department, UCSD, Spring 2012.

RESOURCES: CREATING AND PUBLISHING VISUALIZATIONS OF CULTURAL DATA

analyze large image and video collections (selected image processing software):

batch image processing tools from Software Studies Initiative

shotdetect

Visualize large image and video collections:

Visualizing Image Sequences with ImageJ (Montage, Slice, scaling, etc.)

Preparing data with UNIX commands:

basic UNIX for working with data

introduction to Unix commands for humanities students (Scholars Lab, University of Virginia)

Summarizing data in Excel:

Guide to data cleaning in Excel

5 methods to summarize data in Excel

How to consolidate data in Excel and graph it

techniques to add and count Excel data

Add numbers based on multiple conditions in Excel

Network analysis and visualization for humanities:

Demystifying Networks

Design your visualization to look nice and make impact:

Graphic design principles for Information visualization

Inspiration:

http://tulpinspiration.tumblr.com/

http://infosthetics.com/

Data and visualization blogs worth following

popular websites and blogs about visualization

Innovative visualizations of temporal flows

Use of visualization in museum web site and online media collections

Visualization design patterns:

InfoVis wiki list of visualization design patterns

visualizing cultural data - my favorite tools:

1) Excel

2) Mondrian

3) ImagePlot

lists of other visualization tools:

datavisualization.ch list of visualization and mapping tools

.net list of the top 20 data visualization tools

WikiVis list of visualization tools

Popular software tools and applications for creating visualizations

Sound visualization software

Software to analyzing and presenting online digital collections

Over 100 Incredible Infographic Tools and Resources (Categorized)

Other guides to digital humanities tools:

Tooling Up for Digital Humanities (Stanford University)

Digital Research Tools

Graduate students: academic networks and publishing / citation platforms:

Open access humanities journals (Wikipedia)

Directory of open access academic journals

Mendeley

Research Gate

Academia.edu

Undergraduate students: promoting and presenting your work:

Online portfolio preparation / graduate schools applications

UCSD
Spring 2012
Visual Arts Department. UCSD

undergraduate course: VIS 149 / ICAM 130: Special Topics
(topics only for undergraduates are marked in **pink**)

graduate course: VIS 219: Special Topics
(additional readings and topics for graduate class are marked in **purple**)

COURSE INFORMATION

Instructor: Lev Manovich
www.manovich.net
www.softwarestudies.com
e-mail: manovich@ucsd.edu
Office hours: Thursday 2:00-3:00, Cafe Roma (or by appt.)

VIS 149 / ICAM 130
Th 3:30-6:20pm, VAF 228

VIS 219
W 3-5:50pm, VAF 366

COURSE DESCRIPTION

Data Visualization and Computational Art History

"The next big idea in language, history and the arts? Data."
New York Times, November 16, 2010.

This class covers basic techniques of data visualization and "media visualization" techniques (visualizing large image collections). We explore how these techniques can be used together with digital image processing to visualize patterns in large cultural visual data.

Throughout the course, we use a single large data set (1,000,000 images and user data from deviantArt social network for user-generated art). We will also consider museum collections data which recently became available via museum APIs.

Students work in groups on final projects. Succesful final projects become parts of our collective paper about deviantArt.

Course goals:
1) learn visualization and data exploration skills;
2) learn how to revise and refine visualizations, interpet them and write about them;
3) create original collaborative visualization / digital humanities projects;
4) become familiar with the relevant material in media art, digital art, graphic design, artistic visualization, social computing, digital humanities;
5) discuss theoretical questions related to using visualization and computer data analysis as a methodology for art history, visual culture, communication, media, new media, and game studies.

Software:

Excel, Mondrian, manyeyes, R project, ImageJ, and ImageJ macros developed by Software Studies Initiative at UCSD).

We will also have access to state-of-the- art supervisualization systems at Calit2 (HIPerSpace) for presentation of final projects.

COURSE OUTLINE

spring 2011

(topics only for undergraduates are marked in **pink**)
(additional readings and topics for graduate class are marked in **purple**)

class 1
overview of course content;
data visualization and digital humanities

class 2
visualizing one dimensional and two dimensional data: bar chart, pie chart, line graph, scatter plot (Excel);
histograms (Mondrian)
data transformations (log, etc. - Mondrian, Excel)
brief introductions to:
spatial visualization, text visualization, network visualization (manyeyes)
- assign homework 1

class 3
generating data summaries; organizing and cleaning up data; spreadsheet formulas; random sampling (Excel, Google Docs)
graphic design principles for information visualization
- review homework 1 v.1

class 4:

generating data summaries, continued;
visualizing multi-dimensional data: scatterplot matrix, radar and spider plots, mosaic plot, small multiples (Excel, Mondrian, Google Docs):
- revised homework 1 v2 due;
- assign homework 2;

class 5:
basic descriptive statistics;
use of visualization in museum and collection interfaces;
museum collections APIs;
visualizations of temporal processes

optional workshop for graduate students: working with digital images

class 6:
media visualization techniques (ImagePlot);
exploring large cultural data sets: summary statistics vs. information visualization vs. media visualization
features and feature space
media analytics
visualization and cultural categories (style, period, etc.)

class 7:
media visualization techniques (ImageJ, custom-plug-ins) : montage, slice, image average;
batch processing: scaling, cropping, padding (ImageJ)
relevant projects from media art
working with images collections: file formats, compression options, color spaces, image histograms
taxonomy, folksonomy, metadata, digital libraries
- start final class project

class 8:
basic data analysis with R;
basic UNIX commands for working with data;
preparing visualization projects for the web; using social media for sharing projects
review blogs and projects of leading visualization designers
visualization and theory in humanities and social science - Bruno Latour's arguments
- homework 2 v1. due;
- work on final project

class 9:
analyzing multi-dimensional data with R
spatial history, network analysis
preparing your portfolio;
(software studies: thinking theoretically about software tools)
homework 2 v2. due
- work on final project

class 10:
computer science research of online social networks (social computing);
big data paradigm in humanities and social sciences - possibilities and questions
digital humanities and computational social science: questions, current work, funding
(if time allows: Franko Moretti: Abstract Models for Literary History)
- review final project v.1

RELATED COURSES

145B: image analysis and visualization techniques for digital humanities (Manovich, UCSD, spring 2011)

digital humanities++ (Manovich, UCSD, spring 2011)

MAT259: Visualizing Information (Legrady, UCSB)

CS 171: Visualization (Pfister, Harvard)

CS4488: Data Visualization (Jeffrey Heer, Stanford)

COURSE REQUIREMENTS AND POLICIES

Attendance

You are allowed to miss one class meeting without an excuse.

Any additional absence without a proper excuse (doctor's notice) will lower your final grade by half a letter grade (for example, B becomes B-.)

Grading:

Course grading:
A: 90-100 points
B: 80-89 points
C: 70-79 points
D: 60-69 points
F: 0-59 points

Grade breakdown:
Homework 1: %20
Homework 2: %20
Final project v.1 : %30 (due in class 10)
Final project v.2: %30 (due at the end of final week)

Homework grading:

I will grade v2. of each homework only:

A (20 points) - Your visualizations are done correctly (clearly show patterns in the data, easy to read, have all needed labels, follow contemporary design principles)
B (15 points): small problems
C (10 points): major problems

If you don't have homework v.1 on the due date, I will take 10 points off.

Grading of the final project:

Similar to homework grading but I also include assesment of your writing in the grade. You need to clearly describe the data used, the methods, and explain and interpet each visualization.

The grade for the final project will consist from two parts:
1) grade for v. 1 due in class 10;

2) grade for version 2 due at the end of the final week.

Final project details:

Final Projects due: 11:55pm, Saturday, June 16.

Students will work on final projects in groups.

Each group final project should use the techniques learned in the class to explore some aspects of our DeviantArt (dA) dataset which consists from 1 million images, their metadata and demographics info on their creators.

Images, Metadata, Features:

DeviantArt Traditional Art and Digital Art: images, metadata and selected features:
images and data

Explanation of user symbols in the data files

Note: the data files we prepared for you only contain selected features from the complete set of 400 features we extracted from each image. if you want, we can make these additional features available.

explanation of features
Features documentation - Google doc

Note that you can also run QTIP image processing program which generates 32 bin brightness histogram. Most importantly, the program also outputs a number which tells you if a given image is color or black and white - so you can use to separate these two types of images.

A simple way to get started:

Identify one (or more) groups of similar images (in terms of content and/or their visual language).

If we only had a thousand images, that will be simple.
The challenge is how to do this in a set of 280,000 images
(Traditional and Digital Art categories)?

You can use the metadata, image features, and various visualization techniques and tools we learned in the class. For example:

- visualize images in different subcategories and see if any of these subcategories contain similar images. Maybe not - but if you find that at least some of the images in a particular category are similar, that's already a start.

- select users who have big image collections and visualize these collections. This can be another way to locate subsets of similar images.

- yet another way is to simply select a few similar images, look at their feature values, and then filter data file (In Excel) to isolate other images which have similar feature values

Of course these techniques can be used together, and you can come up with more tricks.

Once you locate nice sets of similar images, you can do many things, for instance:
- use image features to visualize these sets;
- see how these pools of similar images are distributed in terms of user demographics and countries, and compare the subsets

Other ideas to explore:

Try to find subsets of images which are similar visually and/or thematically.
Do these subsets correspond to the deviantArt subcategories, or do they cut across categories?

You can compare images of some selected members.

Visualize all images in selected categories (you can also use smaller random samples), so the images are which have similar visual attributes or/or content appear close to each in a visualization.

Compare and contrast images in similarly named subcategories (i.e. "drawings") in Traditional

Art and Digital Art categories. Can we see distinct traces of the use of digital authoring tools? (Even if you only isolate subsets of images which show such traces, thats already will be useful.)

If you find significant number of similar images (which may or may not belong to the same subcategories) created by users with similar demographic profiles, you can then discuss these subsets. (For example, what are the differences between traditional paintings by users from country X and country Y?)

The ultimate dream goal of our deviantArt project is to find what people imagine and create today around the world, as reflected in deviantArt sample. Can we isolate all different types of content and visual strategies in this sample? While I dont expect that we can reach this goal in this class (or maybe ever), lets see if we can make some progress.

Another dream goal is to be able to identify and analyze cultural influence. Can we automatically find in our sample the cases where a new technique, style or content introduced by an influential user (a user who is followed by many other users) later appear in images by users who follow him/her? If we analyze a large number of such confirmed cases of influence, what does this tells about how influence and imitation work in general? If you want to try to work on this, we can make available to you additional data files we have which show who follows who (due to the size of these files, you should only attempt this if you know mySQL or can manipulate big data via programming.)

Data set:

Background info:

DeviantArt

Wikipedia article on deviantArt

Alexa.com infomation for DeviantArt web site

Our data set currently consists from from 1 million images, their metadata and demographics info on their creators.

DeviantArt image galleries are organized by categories (when a user uploads an image, s/he selects the category for this image).

Examples of categories:

http://browse.deviantart.com/photography/macro/

http://browse.deviantart.com/traditional/paintings/
http://browse.deviantart.com/manga/digital/vector/

Some categories indicate media type - but this does not mean that all people who mark their submitted images with this category using it correctly. So, for instance, you may find drawings in "painting" category, and paintings in "drawing category."

Your visualizations do not need to use all images - you can use random samples, or show subsets of images which have some connections to each other.

How to collaborate using Dropbox:

Create a folder for your group final project inside **2012 Spring winter final projects** Dropbox folder.

Name your project folder as follows:
DeviantArt student_lastname1 student_lastname_2 student_lastname_3 ..

for example:
DeviantArt Jones Kim Beck

Final project format:

The project should be formatted as **Word .doc**, or PDF or a Powerpoint.

See our projects (Software Studies Initiative projects) and papers (Software Studies publications) for examples of how to organize writing and visualizations together.

Text length:

undergraduates: around 1000 words (longer is OK).
graduates: around 1500 words (longer is OK).

Insert small size versions of your visualizations (saved as jpeg or png) into the doc.

I will give you a template to follow (text width, font, headings size, etc.)

Keep the size of your doc **below 5 MB**.

PDF option:

you can create a high resolution visualization poster in any software (Photoshop, Illustrator, etc.) and save it as PDF.
In this case, insert you high-res visualizations into your design.
Keep the PDF below 100 MB.
Also, save your design as JPEG file.

You also needs to submit all high res visualizations and the data sets used in the analysis.

Indicate the roles and the contributions of each group member in the .doc

Make sure labels and text is readable - see this example of a complex but readable visualization

Publication/Promotion:

if you final project is succesful and if you permit this, we will feature it on softwarestudies.com and promote it.

Final files to submit when the project is complete:

In your final project Dropbox folder:

move all files you created earlier which are not used in the final project submission into their own subfolder and call it <other>

Your final submission should contain the following files:

1) document saved it as Word .doc file.

name it:

Final_youlastname.doc (or Final_youlastname.pdf, or Final_youlastname.jpg.)

Make sure this document contains
- **the names of all the students in the group**
- **description of the contributions of each student**

- descriptions of the parts of the data set used in all visualizations / analysis
- labels for each visualization indicating data used and what is shown on X and Y axis.

2) A subfolder named **Scaled_visualizations**.
Place scaled down versions of the visualizations and any other illustrations you are using in your project in this subfolder. (If you are using HTML format for your essay, link to these images. If you are making .doc document, you can directly paste these images into the doc.)

3) A subfolder named **Full_size_visualizations**. Place full-size visualizations in this subfolder. They should have exactly the same names as the scaled down versions - however, for scaled versions, add "scaled" to their names; for full versions, add "full".

For instance:

scaled down version:

New_York_Tribune.1886_1886.Montage.rows_40_columns_40.w900_h600.scaled.jpg

full version:

New_York_Tribune.1886_1886.Montage.rows_40_columns_40.w12000_h8000.full.jpg

4) A subfolder named **Data**. If you did any measurements of the images, or added new metadata and used them in visualizations, put these measurement files in this subfolder. Make sure that the data files have descriptive names, and also a label row with clear descriptions of all columns.

5) Optional:
If you also prepared infovis poster, create subfolder **Poster** and put .jpg and pdf. files of your poster there.

All files and folders have to named consistently and clearly. Points will be taken off for careless naming of the files.

Points will be also taken off if your graphs don't have clear labels.

Academic Integrity

Integrity of scholarship is essential for an academic community. The University expects that

both faculty and students will honor this principle and in so doing protect the validity of University intellectual work. For students, this means that all academic work will be done by the individual to whom it is assigned, without unauthorized aid of any kind.

In this course, we expect that all assigned practical assignment are done by students individually. Group projects, on the other hand, are designed to be done collaboratively.

Students with Disabilities

Students requesting accommodations and services due to a disability for this course need to provide a current Authorization for Accommodation (AFA) letter issued by the Office for Students with Disabilities (OSD), prior to eligibility for requests. Receipt of AFAs in advance is necessary for appropriate planning for the provision of reasonable accommodations. OSD Academic Liaisons also need to receive current AFA letters.

For additional information, contact the Office for Students with Disabilities:
858.534.4382 (V)
858.534.9709 (TTY) - Reserved for people who are deaf or hard of hearing
osd@ucsd.edu
http://disabilities.ucsd.edu

REQUIRED READINGS

All required readings will be available online at no charge.

Popular software tools and applications for creating visualizations:

(This list is based from Jim Hollan USCC information visualization seminar, with my additions)

ChronoViz (http://www.chronoviz.com): ChronoViz is a tool to aid visualization and analysis of multimodal sets of time-coded information, with a focus on the analysis of video in combination with other data sources.

D3: Data Driven Documents (http://vis.stanford.edu/papers/d3): Data-Driven Documents (D3) is a novel representation-transparent approach to visualization for the web.

Ggobi (http://www.ggobi.org/): Ggobi is an older open source visualization program for exploring high-dimensional data.

Google Visualization (http://code.google.com/apis/visualization): API and Charts.
Many charts can be generated by entering a single URL string.
You can also interactive chart wizard.

ImagePlot (http://lab.softwarestudies.com/p/imageplot.html): a unique visualization tool for exploring patterns in image collections.

ManyEyes (http://manyeyes.alphaworks.ibm.com): Web-based visualizations.

Mondrian (http://www.theusrus.de/Mondrian/)**:** An application for interactive exploration of multi-dimensional data sets.

Matlab (http://www.mathworks.com/products/matlab/description4.html): Used primarily for scientific and engineering visualization but a real advantage if you use Matlab for data analysis.

NodeXL (http://nodexl.codeplex.com/): Popular network visualization software which works with Excel spreadsheets. Currently Windows only. Network graphs using edge and vertex lists stored in an Excel 2007 or Excel 2010 workbook.

Prefuse (http://prefuse.org/): Java API for information visualization.

Prefuse Flare (http://flare.prefuse.org/): ActionScript 3 library for data visualization in the Adobe Flash Player.

Processing (http://processing.org/): Popular language and IDE for graphics and interaction.

Protovis (http://vis.stanford.edu/protovis/): JavaScript tool for Web-based visualization.

Timeline (http://timeline.verite.co/): create multimedia timelines.

VTK (http://vtk.org/): Library for 3D and scientific visualization.

Sound visualization software

sonicvisualiser

COURSE SCHEDULE

class 1:
overview of course content;
data visualization and digital humanities

references:

The program of Museum and the Web 2012 San Diego conference:
http://www.museumsandtheweb.com/mw2012/about

class topics:

slides of my lecture

homework for class 2:

Install on your laptop following software (make sure to bring your laptop to every class):

1) Mondrian software

2) Excel (if you don't have already on your laptop, download 60 day trial version).

Read:

1) http://lab.softwarestudies.com/2008/09/cultural-analytics.html

2) Manuel Lima. Outburst of Visualization (2010). In Lev Manovich, Jeremy Douglass, William Huber, *Mapping Time*. gallery@calit2, forthcoming 2012.
Visit the web sites and projects referred in the article.

3) In 500 Billion Words, New Window on Culture. NYT, 12/16/2010.
Visit http://ngrams.googlelabs.com.
(more information: http://www.culturomics.org/)

5) for graduate students:
Introduction to Digital Humanities Journal

optional:

Lev Manovich. "Cultural Analytics: Visualizing Cultural Patterns in the Era of 'More Media.'" DOMUS, Spring 2009.

Start exploring the leading infovis web galleries:

infosthetics.com
visualcomplexity.com

---

class 2:
visualizing data which has one or two dimensions / visualizing using maps / basic text visualization

software:
Excel
Mondrian
manyeyes
Google Chart Tools
(optional: Google spreadsheets (part of Docs), OpenOffice, TextWrangler, Tableau)

practice data sets:
van_gogh_data.txt
Time magazine covers tags (1923-1990)

references:
history of data visualization
Wikipedia article on visualization techniques

examples of interesting map visualizations: http://pinterest.com/shashashasha/pins/

Data and Image Models
Wikipedia article on GIS

Google Chart Tools
Google Visualization Playground
Google Chart Wizard

class topics:

visualizing data which has one dimension:
pie chart, bar chart (Excel),
histogram (Mondrian) (explain differences between bar chart and histograms)

visualizing time series: line graph (Excel, Mondrian)

visualizing data which has two dimensions: scatter plot (explain differences between line graph and scatter plot) (Excel, Mondrian)

data transformations (log, etc. - Excel graphs axis options; Mondrian)

visualization using maps (manyeyes)

basic text visualization (manyeyes)

visualization using web services (Google charts API)

homework for class 3:

practical assignment 1 - version 1 due before class 3:

use basic visualization techniques to explore interesting cultural patterns contained in the list of images (and their user data) in deviantArt Traditional art category - from our deviantArt one million images download.

DATA:
dA.traditional_art.images_metadata.txt (it may take a little time to download, be patient).

(if this is too simple for you, let me know and I will give the data for the complete set of

one million images in our sample.)

You can use of any of basic techniques for visualizing one dimensional data (pie char, bar graph, histogram) and two dimensional data (scatter plot) covered in class 2. You can invent new visualization techniques which build on these basic techniques, or use network visualization or spatial visualization.

You need to produce minimum TWO different visualizations.

Save your visualizations as .PNG.

Label each axis and add a **descriptive caption** which makes your visualization self-explanatory.

GOALS:

1) Get experience with basic visualization techniques for 1D and 2D data (i.e., visualizing one or two columns of data);

2) Practice exploring a new cultural data sets using visualization with the goal of finding interesting patterns;

3) Practice using principles of contemporary graphic design to produce clear and effective visualizations.

HOW TO NAME THE FILES:

YourLastName.hw1.vis1.png
YourLastName.hw1.vis2.png

NOTE: following feedback received in class, you will have another week to refine or redo your visualizations.

HOW TO SUBMIT:

Use Dropbox folder for the course - it is “2012 Spring VIS 219” (for graduate students) or “2012 Spring ICAM 130 and VIS 149” (for undergraduate students). Create a folder inside a course folder and name it using your last name. Copy the homework files into this folder.

READ:

1) explanation of basic graph types (from manyeyes):

bar chart
line graph
scatter plot

optional - graph types specific to manyeyes:

phrase net
network diagram
world map

2) guide to graphic design for information visualization

VIEW:

Hans Rossling 2006 lecture (video on TED)
Aaron Koblin 2011 lecture (video on TED)

---

class 3:
organizing, summarizing and cleaning data; using spreadsheet formulas; data sampling

software:
Excel, Mondrian, Google Docs
(optional: Google Spreadsheet, Google Refine, Fusion Table)

summarizing data in Excel - guides:

Guide to data cleaning in Excel

5 methods to summarize data in Excel

How to consolidate data in Excel and graph it

techniques to add and count Excel data

Add numbers based on multiple conditions in Excel

NOTE:
ALL FILES USED IN CLASS DEMOS ARE AVAILABLE IN CLASS DROPBOX FOLDER "DEMO FILES"

class topics:

review practical assignment 1 (v. 1.)

Calculating and graphing data averages:

using subtotals command in Excel to summarize the data:

1) sort the column you want to summarize

2) use Subtotal command on this column

3) copy the results:

how to copy subtotals results

How to consolidate data in Excel and graph it

common data formats for visualization:
text document, network data format, table (XML, spatial data, database, etc.)

common formats for storing data tables:
spreadsheet, tab-delimited text file, comma separated text file

converting between file formats (xls, .csv, .txt)

data types: integers, floating point numbers, strings

working with spreadsheet data using functions

using a function in Google spreadsheet to convert a table from a web site into Google spreadsheet

converting web page tables into Google spreadsheets

example:
a list of van Gogh paintings from Wikipedia page
cleaned data using Google spreadsheets functions
the titles (with periods added) data uploaded to manyeyes

Creating a random data sample in Excel:

add a new column to your spreadsheet;
fill it with rand formula:

```
=rand()
```

copy this column into a new column using “paste special” (select “copy values” option)
sort spreadsheet by the new column
select first N rows

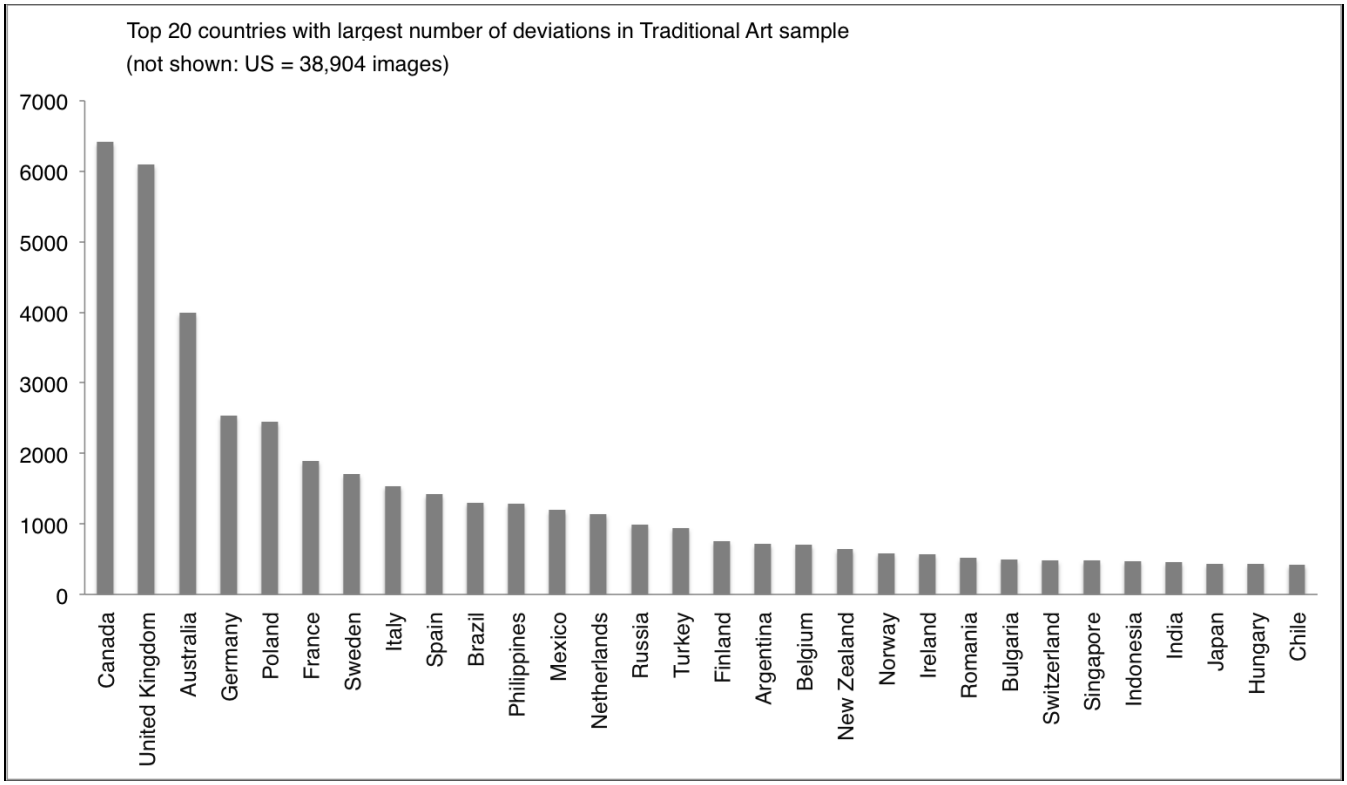
Homework for class 4:

1) Revise / finish homework 1 and put the new finals into the class Dropbox.

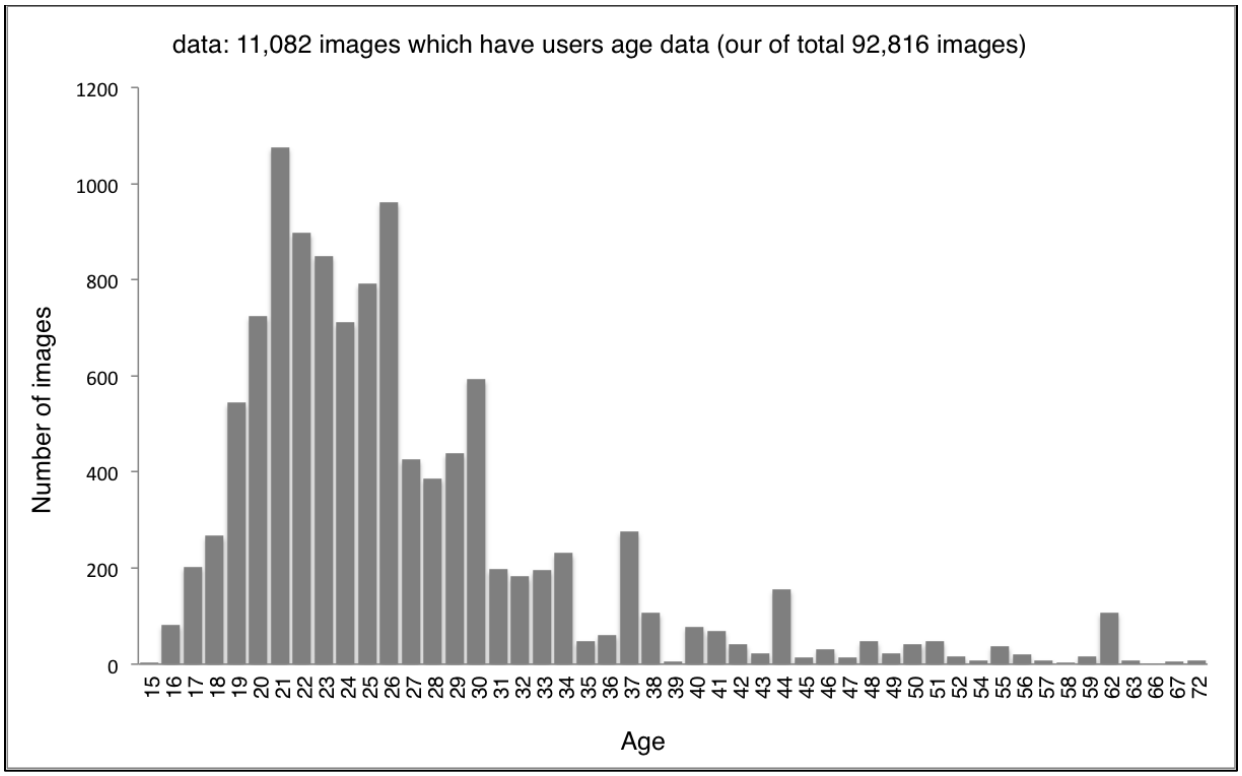
Name the submitted files as follows:

YourLastName.hw1.vis1.final.png
YourLastName.hw1.vis2.final.png

Example of a visualizations (created by me) which meet the homework requirements:



Steps to produce this chart in Excel:
-sorted the data using country column;
- used subtotals command to generate the he list of images per country;
- using how to copy subtotals results technique, copied this data into a new worksheet;
- plotted the summary as a barchar, and adjusted graph options.)



To produce this chart, I followed same steps - but used subtotals command with Age column

Notes for the homework visualizations:

- make sure you laxel axis and have a title/explanation of what is being shown;

- work on graphic design of your graphs - dont just settle for the default colors, line widths, font sizes, etc. which the program produces. I did these adjustments in Excel but you can also bring your graph into Illustrator/Photoshop to make graphic changes.

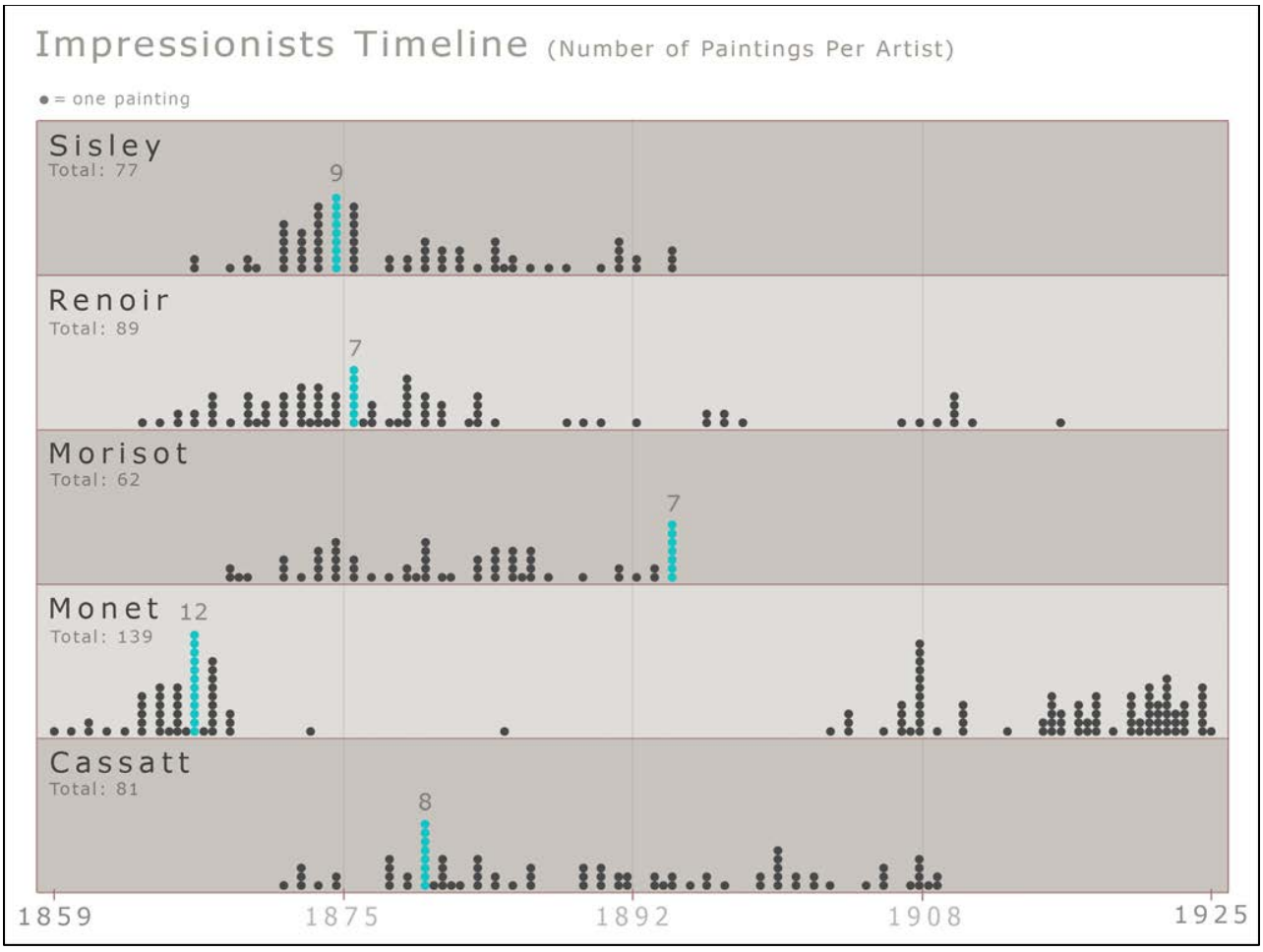
- dont just graph anything - pick up parts of the data which show interesting patterns

- graphs you may want to try do (more challenging then my examples): 1) graph the system of categories showing their relationships and numbers of images and/or users for each category; 2) graph the relations betwen age, sex, and country dimensions

- learn how to use Subtotals command in Excel - I used to produce the sample graphs

- review my Graphic design principles for Information visualization
and visit examples of great visualization sites linked there

- Here is an example of a good visualization (the numbers of paintings produced by a few Impressionist artists per year - only partial data is shown) - follows design principles, explicit labels, shows lots of data in a single visualization. Created by Megan O'Rourke (UCSD undegrad) in my winter 2012 class:



the larger version of this visualization

2) How visualization designers work ? Review descriptions of some of the projects from datavisualization.ch (feel free to look at other projects as well):

How We Visualized 23 Years of Geo Bee Contests
Visualizing The World's Well-Being

How We Visualized America's Food and Drink Spending
Visualizing The Health Care Reform

---

class 4:
visualizing multi-dimensional data: radar plot, small multiples, mosaic plots, scatterplot matrix (Excel, Mondrian, Google Docs);
discussion: use of visualization in museum and collection interfaces;
museum collections APIs;
web scraping

references:
Multi-dimensional visualization lecture

NOTE:
ALL FILES USED IN CLASS DEMOS ARE AVAILABLE IN CLASS DROPBOX FOLDER "DEMO FILES"

class topics:

visualizing multi-dimensional data:

radar plot (Excel)

small multiples (Gauguin)

mosaic plots (Mondrian)
explanation of mosaic plots

scatterplot matrix (Mondrian)

Homework for class 5:

1) Review techniques for summarizing data in Excel and practice using homework

data set:

using subtotals command in Excel to summarize the data:
how to copy subtotals results

How to consolidate data in Excel and graph it

additional guides - use as needed:

Guide to data cleaning in Excel

5 methods to summarize data in Excel

How to consolidate data in Excel and graph it

techniques to add and count Excel data

Add numbers based on multiple conditions in Excel

2) Review the examples of museum web sites using visualization and prepare to discuss them in class

3) View more museums put their collections online - visit linked museum sites

4) Check the examples of large online collections of cultural data / digital archives:

Europeana
Internet Archive
Chronicling America: Historic American Newspapers

optional:
Digital Public Library of America
Hathi Trust Digital Library
Google Books
archive list from Digging Into Data competition

5) Check the examples of innovative visualizations of temporal processes and prepare to discuss them in class

6) You next practical homework - due in class 8:

Choose between these two options:

Option A:

Create a visualization poster or a mini-essay exploring interesting patterns in Cooper Hewitt Museum collection data Cooper-Hewitt-objects.csv
(this file is in DEMO FILES Dropbox folder).

Information about this collection data:

http://www.cooperhewitt.org/collections/data

Homework format options:

- a high resolution poster (JPEG) containing a few visualizations with text explanations. Name the file: Cooper-Hewitt-yourlastname-poster.jpg.

Maximum file size: 10 MB.

- Word .doc file containing visualizations and text (maximum 500 words). Name the file:
Cooper-Hewitt-yourlastname-poster.jpg.
Maximum file size: 10 MB.
If you choose this option, you can also submit high resolution versions of any of the visualizations along with the .doc. Name them as follows: Cooper-Hewitt-yourlastname-vis1.jpg, Cooper-Hewitt-yourlastname-vis2.pg, etc.

Option B:

use media visualization and information visualization to explore "similarity space" of a number of Impressionist artists

Details:

Using our dataset of selected paintings by Impressionist artists and the metadata about these paintings, explore the similarities and differences between their art and careers:

images: Impressionism.zip

URL: http://neen.ucsd.edu/Impressionism/

Data: two files located inside DEMO FILES directory in class Dropbox folder:
Impressionism.all_images.txt (metadata for all images with dates)
Impressionism.images_with_dates.txt (metadata for all images with dates)

Your visualizations should explore similarities/differences between:

artists's careers
periods in their careers
all paintings of each artist
groups of works
individual works

You don't have to use all these dimensions - but do use **at least two.**

You can supplement our existing collection by external metadata (for example, artists biographies, more complete lists of their works), and also additional images of their works not in our collection.

Your visualizations can use images, their automatically extracted features, their metadata (titles, dates, dimensions), and other metadata about artists' careers.

Your visualization poster should include at least one **image plot,** at least one other type of **media visualizations** (montage, orthogonal view, and/or average), at at least one i**nformation visualizations** which use standard (bar charts, scatter plots, line graphs, etc.) or unique techniques.

Combine all plots and text into a single visualization poster.

Include **labels** and **text** explaining it (what data are used; legend; and anything add you want to add.)

Make sure that the visualization file you put in the Dropbox is below **10MB**.

Examples of the successful student work from a previous class:

single visualization:
http://lab.softwarestudies.com/2012/04/visualizations-of-impressionist-artists.html

final posters:
http://lab.softwarestudies.com/2012/04/impressionism-visualizations-final.html

class 5:
basic descriptive statistics;
discussion: use of visualization in museum and collection interfaces;
museum collections APIs;
sources of cultural data;
visualizations of temporal processes

references:

wikipedia page on summary statistics

free statistical web tools: www.wessa.net

class topics:

basic descriptive statistics: mean, median, standard deviation (Excel)

Use of visualization in museum and collection interfaces

sources of cultural data;

Introduction to cultural collections APIs:

list of museum and collections APIs
Brooklyn museum API
example using Brooklyn museum API
introduction to using APIs

Copper-Hewitt collection data

visualizations of temporal processes

Introduction to web scraping:

web scraping with SiteSucker

resources:

Scraping for Journalism: A Guide for Collecting Data
Data Driven journalism (Wikipedia)

Homework for class 6:

1. Learn ImagePlot software:

   1) go to ImagePlot web page.

   2) download ImagePlot.zip

   3) go through ImagePlot documentation (including all tutorials)

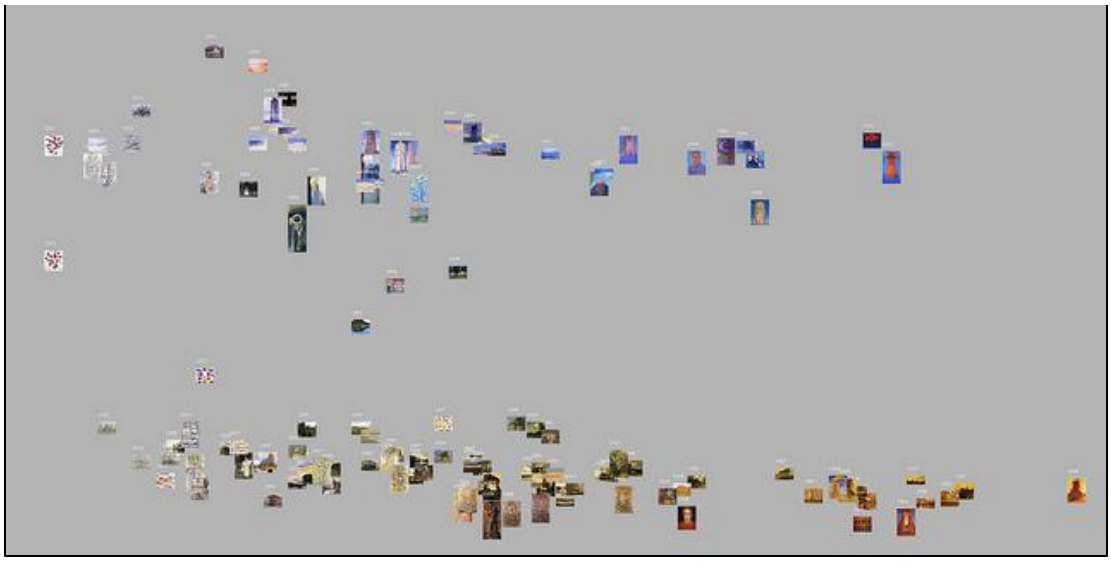
   4) Create the following visualization using ImagePlot and place it into your class Dropbox folder:

   -using Mondrian images and data provided with ImagePlot, create a visualization of all the images organized by median saturation (X-axis) and median hue (Y-axis).
   - Select canvas size, proportions, the size of images and background color which you feel work best to show patterns in this image set.
   - Include labels showing dates in your visualization.

   Name your visualization: YourLastName.Mondrian.jpg

   Keep the size of the visualization below 10 MB.

   Example of the correct visualization (scaled down version - your visualization should be larger):



high resolution version on Flickr

2) read:
Lev Manovich. Style Space (parts 1 - 4), published on softwarestudies.com, 8/-10/2011.

---

optional workshop for graduate students: working with digital images

Contents:

Image formats: JPEG, PNG, TIFF

saving images in JPEG format - options

color spaces: RGB, HSB (illustrated using ImageJ Color3D plug-in)

automatic image correction (Mac Preview, iPhoto)

image histograms (ImageJ)

basic image editing using free software: combine images, cut parts, add titles and other text (Mac Preview, Pixlr)

processing folders of images - format conversion, scaling, cropping,

(using ImageJ)

---

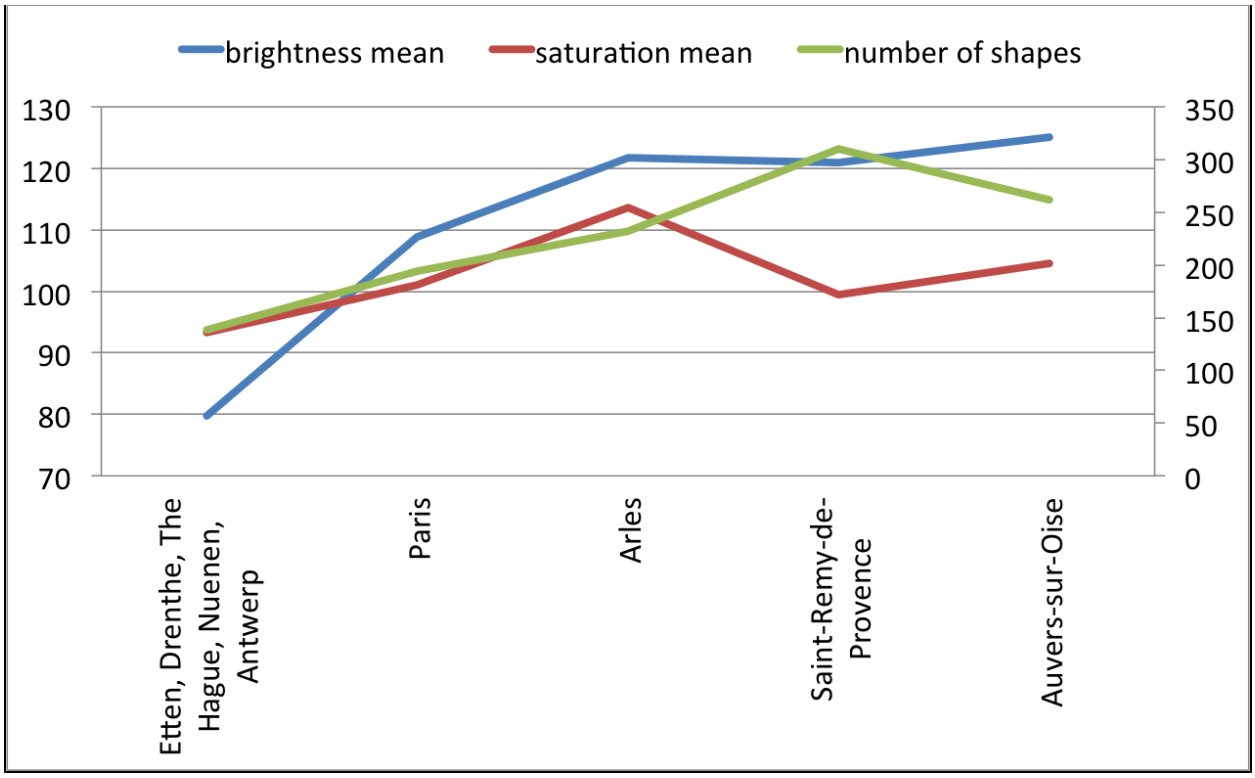
class 6:

Exploring large cultural data sets:
summary statistics / graphs vs.
information visualization vs.
media visualization

summary statistics: Selected features of van Gogh image set summarized using conventional division into periods based on places where the artist lived

place	number of images	dates	brightness mean	saturation mean	number of shapes
Etten, Drenthe, The Hague, Nuenen, Antwerp	196	11/1881 - 4/1886	79.75	93.31	138.4
Paris	199	4/1886 - 3/1888	108.79	101.12	194.17
Arles	161	3/1888 - 4/1889	121.69	113.58	231.71
Saint-Remy-de-Provence	138	5/1889 - 5/1890	120.97	99.5	309.75
Auvers-sur-Oise	81	5/1890 - 7/1890	125.08	104.48	261.39

summary statistics graph: Line graph of selected features of van Gogh image set summarized using conventional division into periods based on places where the artist lived



scatter plot of all van Gogh data points: X-axis = year_month; Y-axis - brightness median

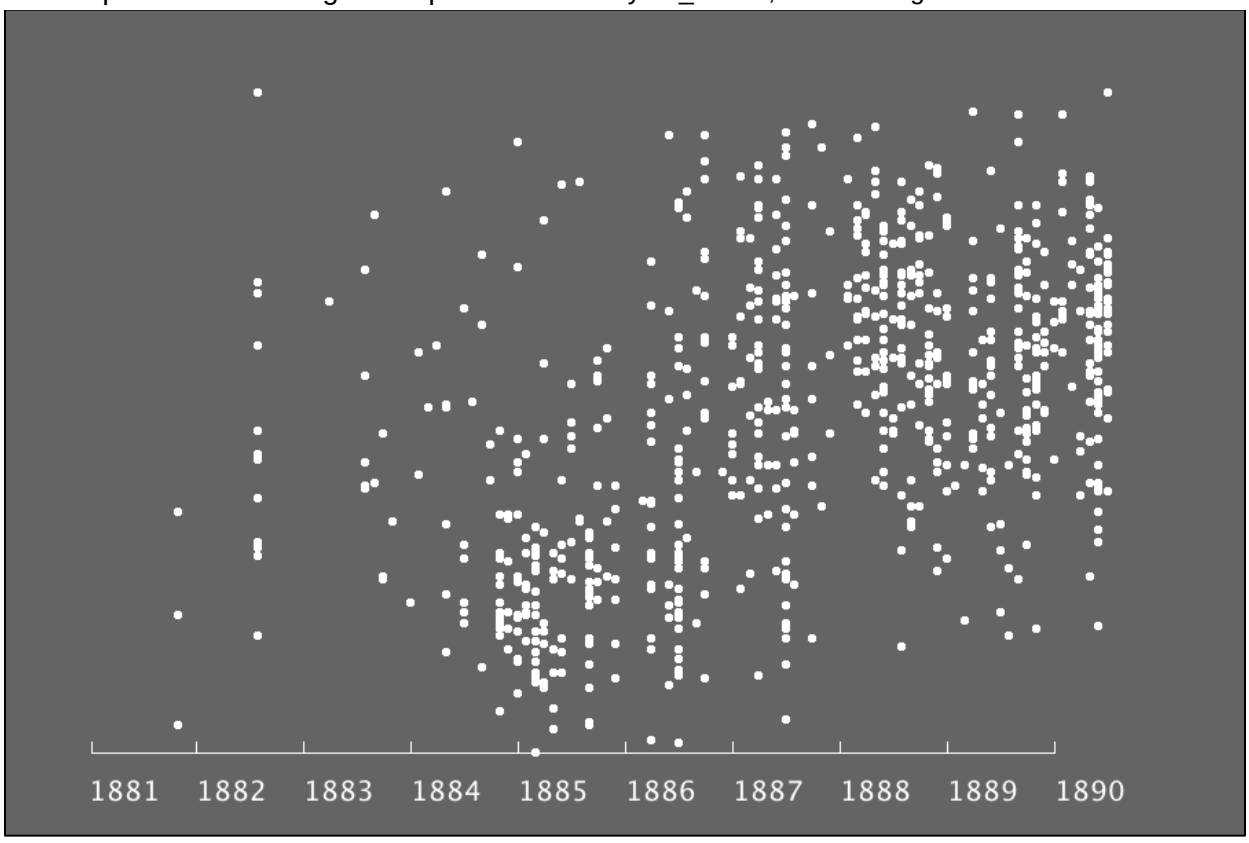
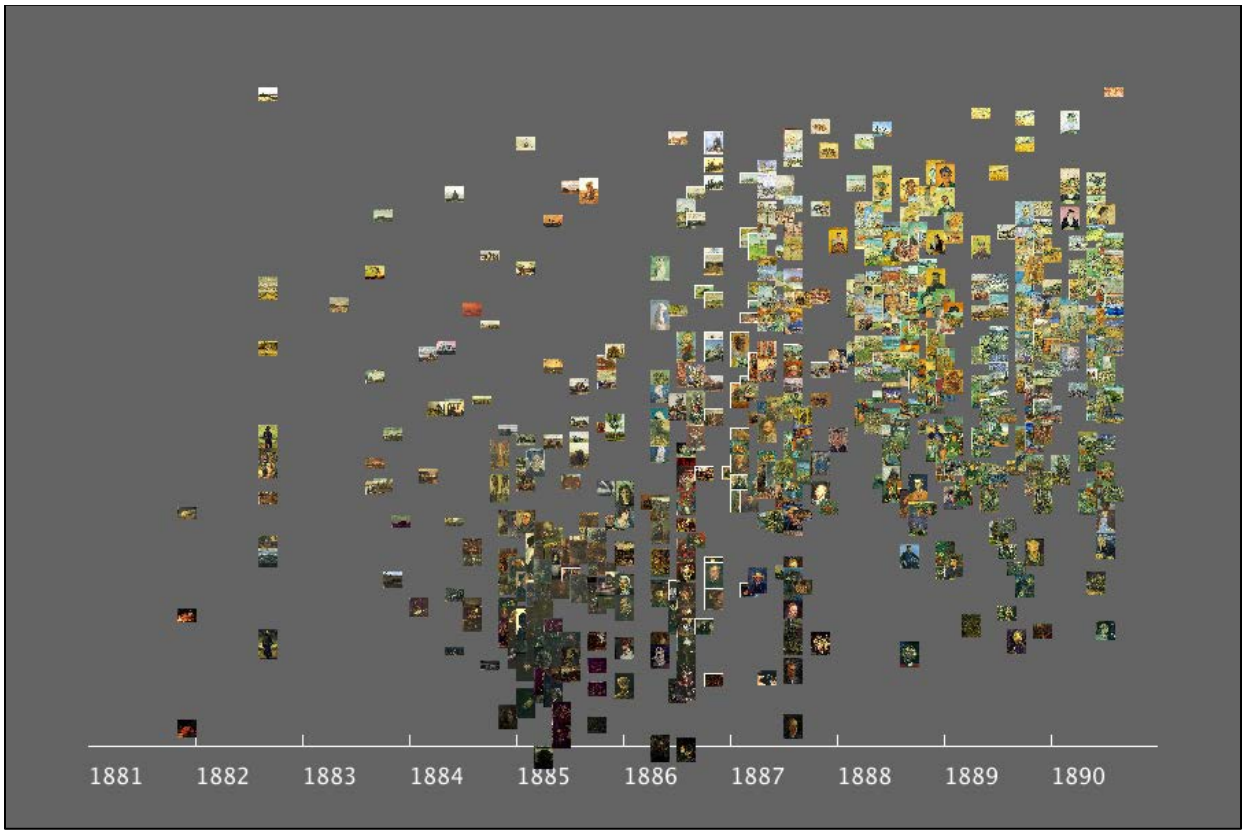


image plot of all van Gogh images: X-axis = year_month; Y-axis - brightness median



examples of image plots: ImagePlot examples gallery

fundamental workflow of media analytics (search engines, content-based image retreival, face recognition, video fingerprinting, recommedation systems, surveillance, text analytics, music information retreival, etc.): 1) feature extraction; 2) analysis

features and feature space

media visualization and cultural categories (style, period, etc.)

media visualization techniques: image plots

ImagePlot software techniques and tips

References:

Ronald A. Fisher. Statistical Methods for Research Workers. 1925.

> “Any investigator who has carried out methodical and extensive observations will probably be familiar with the oppressive necessity of reducing his results to a more convenient bulk. No human mind is capable of grasping in its entirety the meaning of any considerable quantity of numerical data. We want to be able to express all the *relevant* information contained in the mass by means of comparatively few numerical values. This is a purely practical need which the science of statistics is able to some extent to meet. In some cases at any rate it *is* possible to give the whole of the relevant information by means of one or a few values. In all cases, perhaps, it is possible to reduce to a simple numerical form the main issues which the investigator has in view, in so far as the data are competent to throw light on such issues. The number of independent facts supplied by the data is usually far greater than the number of facts sought, and in consequence much of the information supplied by any body of actual data is irrelevant. It is the object of the statistical processes employed in the reduction of data to exclude this irrelevant information, and to isolate the whole of the relevant information contained in the data.”

ImagePlot

ImagePlot examples gallery

Feature extraction (computer vision)

Feature space

Face recognition

Content based image retrieval

Video Fingerprinting

Music information retrieval

Text analytics

Recommended:

Anthony Kenny. *The Computation of Style: an introduction to statistics for students of literature and humanities*. 1982.
(PDF of the selected chapters is in in class Dropbox > TEXTS folder)

Homework for class 7:

1) Go through my guide Visualizing Image Sequences with ImageJ (You can skip parts about video.)

2) Using van Gogh images and data provided with ImagePlot, and **ImageMontage** and **ImageSlice** macros, create the following visualizations:

- Montage visualizations which compare subsets of van Gogh images. Images should be **organized by one or more visual features and/or metadata.**

- Slice visualizations which compares which compare subsets of van Gogh images. Images should be **organized by one or more visual features and/or metadata.**

Your visualizations should support one of the two arguments:
1) There are distinct periods (or distinct groups of works with very different visual characteristics created between 1886 and 1890) in van Gogh paintings;
2) There are no distinct periods (or: or distinct groups of works with very different visual characteristics created between 1886 and 1890) in van Gogh paintings.

(Optional: if this helps your argument, you can also add image plot(s)).

Add labels/legends/text explaining your visualizations and how they support the argument.

Combine all your visualizations into a single poster - name it YourLastName.vanGogh.jpg

Keep the size of the poster **below 5 MB.**
Place the poster in your class Dropbox folder.

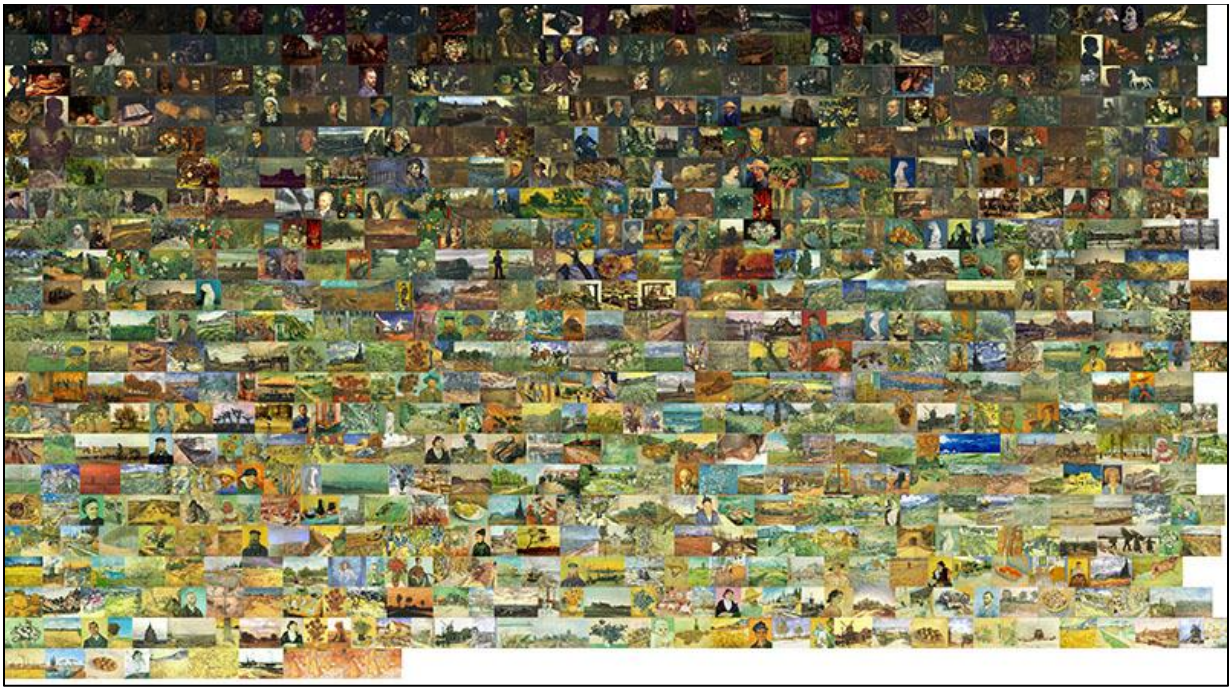
Here are the examples of homework/projects from Winter 2012 which do similar comparisons using ImagePlot and using Impressionist paintings set):

http://lab.softwarestudies.com/2012/04/visualizations-of-impressionist-artists.html

http://lab.softwarestudies.com/2012/04/impressionism-visualizations-final.html

Here are examples of montage and slice visualizations of van Gogh paintings where images are organized by visual features - note that these do not constitute the whole homework since you

have to combine subsets of van Gogh images to each other, and you also need to argue for one or the two possible interpetations of his work. (You can also notice that sorting by median brightness does not work well for images which consist from a very dark and very light areas - a different statistics may work better..)



Montage visualization of 776 van Gogh paintings sorted by median brightness.

Slice visualization of 776 van Gogh paintings sorted by median brightness.

Graduate students:

Read:
Family resemblance
Prototype theory

Daniel Chandler. An Introduction to Genre Theory

Review the following topics:
digital repository
metadata
Europeana (metadata integration)

taxonomy
folksonomy
controlled vocabulary
Folksonomies and Tags

---

class 7:
Artistic projects related to media visualization (creating visualizations out of actual images);
using ImageJ stack commands;
preparing images for visualization - batch scaling, cropping and format conversion;
media visualization using montage, slice, and average techniques;

Note for undergrads: ImageJ is installed on the macs in VAF 228

References:

tutorial covering class topics: **Visualizing Image Sequences with ImageJ**

Lev Manovich. "Media Visualization: Visual Techniques for Exploring Large Media Collections." In Media Studies Futures, ed. Kelly Gates. Blackwell, forthcoming 2012.

Class topics:

(saving images using popular image formats: JPEG, PNG, TIFF;
using JPEG options);

using ImageJ batch commands to measure, re-size and convert image files

working with stacks in ImageJ using built-in commands:

import file sequence into a stack
save stack as an image sequence
virtual stacks
scale stack
crop stack
montage
orphogonal views (“slice” visualization technique)
Z project (“average” visualization technique)

using our custom ImageJ montage and slice macros:

Image_montage
Image_slice

Examples of works which use image averaging technique:

Francis Galton
Jason Salavon

Previous media visualization examples:

Marey
Cinema Redux

Homework for class 8:

1) Prepare homework 2 v1.

2) Start working on final project. ideally, you should form the group, create a group project folder and upload some initial visualizations/sketches/results/ideas. if not, then do some explorations of deviantArt images on your own, and upload your results into your class folder.

Note: you are not expected to explore the whole set of appr. 280,000 images in Traditional Art and Digital Art categories. In fact, you may want to start with the opposite end: identifying "islands" in the "ocean" of all these images which have visual/thematic similarity, or even simply comparing image galleries by different members.

Graduate students:

Read and prepare to discuss:

Bruno Latour, TARDE'S IDEA OF QUANTIFICATION, 2009.

---

class 8:
review homework project 2 v.1
visualizing multivariable data using PCA (Mondrian with R);
work on final projects

preparing visualization projects for the web; using social media for sharing projects

review blogs and projects of leading visualization designers

new http://tulpinspiration.tumblr.com/
new Data and visualization blogs worth following

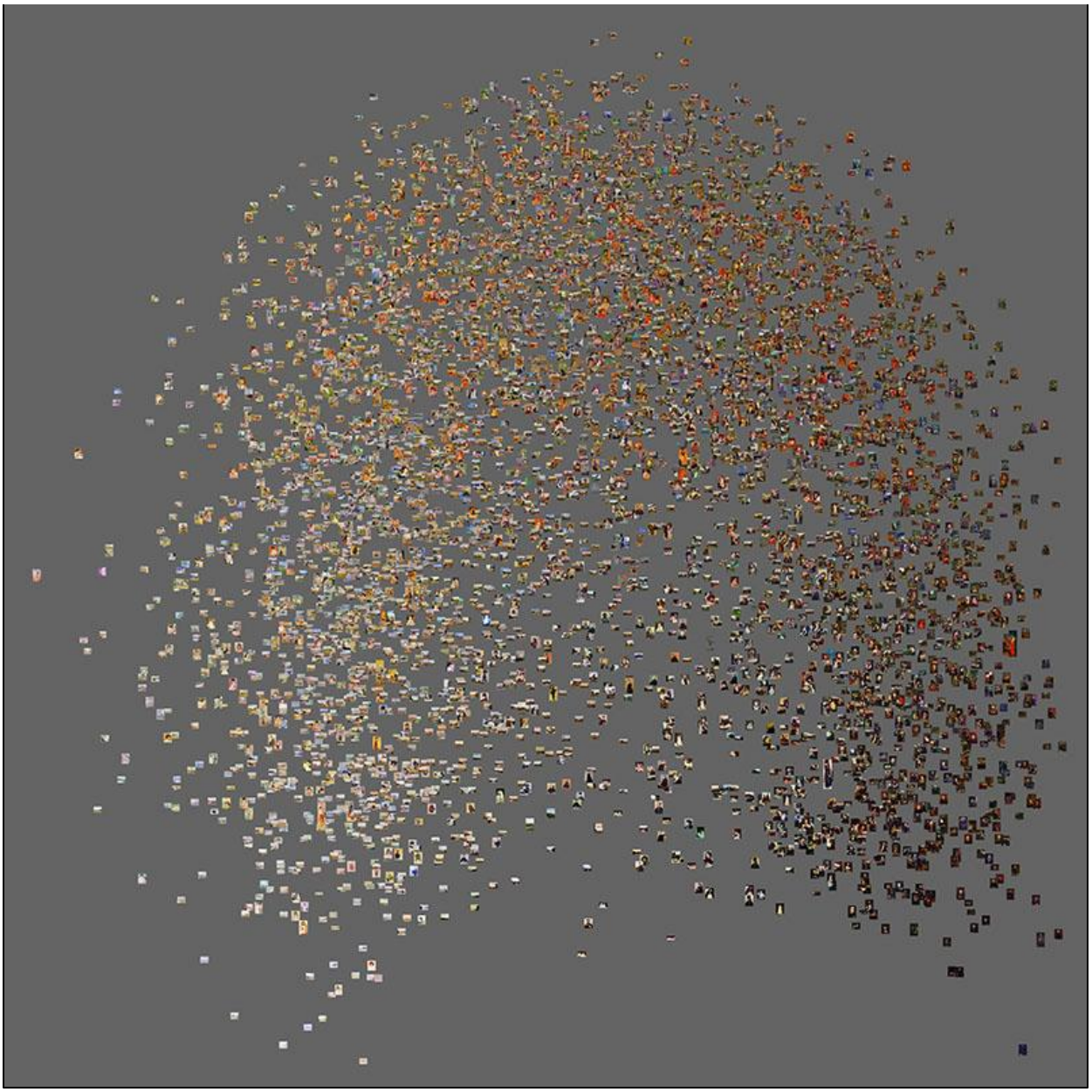
class topics:

introduction to PCA for data visualization

To calculate PCA in Mondrian, you need to install R on your computer, and follow instructions in Mondrian software Help for Starting Rserve

Good guide to using R:
http://www.statmethods.net/

When doing PCA in Mondrian, select Standardize Data (its default).



Visualization of 5433 Impressionist paintings using PCA of 47 lightness features.
X-axis = PCA.1.
Y-axis = PCA.2

Homework for class 9:

1) Finish Homework 2 v2.

2) Work on final project

3) Graduate students:

Go through the following recent examples of quantitative analysis and visualization work in literature studies, film studies and history (you can skip around - just focus on the general ideas, methods and conclusions):

Thomas Elsaesser and Warren Buckland. The Life-Cycle of *Slumdog Millionaire* on the Web (file "Elsaesser_and_Buckland_Datamining_Slumdog.doc" is in TEXTS folder in class Dropbox folder)

Ryan Heuser, Long Le-Khac. A Quantitative Literary History of 2,958 Nineteenth-Century British Novels: The Semantic Cohort Method (Stanford Literary Lab PDF)

Richard White. What is Spatial History?

---

class 9:
review homework project 2 v.2
basic UNIX for working with data;
data analysis with R (review PCA; cluster analysis);
work on final projects

Example of doing PCA with image data using R:

```
impr_with_filename <- read.delim("Impressionism_light.tab.txt")
impr_filename_excluded <- impr_filename[,-1]
impr_pca_cor <- princomp(impr_filename_excluded, cor=TRUE)

 examine results:
plot(impr_pca_cor) // show variance explained by components
plot(impr_pca_cor, type="lines") // same as line graph
biplot(impr_pca_cor) // plots 1st and 2nd components by default
loadings(impr_pca_cor) // show percentage of variables used by each component
```

UNIX for visualization design:

> introduction to Unix commands for humanities students (Scholars Lab, University of Virginia)

basic UNIX commands and skills:

basic commands;
working with directories;

Generating a list of image filepaths with Unix commands
(for creating data files for mageMontage and ImageSlice ImageJ macros, and for ImagePlot if you want to use images across multiple directories)

some useful UNIX commands for working with big data sets:

find size of a directory (in kilobytes):

```
du -sk subdir
```

count number of lines in a file:

```
wc -l filename
```

copy all files from a directory which match some condition
(in this examples: all jpeg images)

```
find . -name *.jpg -exec cp {} /destination/dir \;
```

creating a histogram list of items in a file:

```
sort -n input.txt | uniq -c
```

convert comma-delimited file to a tab-delimited file:

```
cat oldfile.txt | tr '[,]' '[\t]' > newfile.txt
```

delete first column of a file:

```
cut -f2- oldfile > newfile
```

delete 2nd row from a file:

```
sed '2d' file > newfile
```

combine a number of files horizontally:

```
paste file1 file2 file3 > file_combined
```

Homework for class 10:

Read / View:

The Age of Big Data. *New York Times*, February 11, 2012.

example of current big data analysis in social computing:

http://livehoods.org/

example of current big data analysis in digital humanities:

mappingtexts.org (check out http://language.mappingtexts.org/)

examples of companies who do large scale analysis of media and social data:

finlabs.com / article about Blue Fins labs

The Echo Nest

art.sy

Additional readings for graduate students:

read these articles carefully:

Lazer et all. Computational Social Science. *Science*, February 2009.

dana boyd. Six provocations about big data.

browse through these:

Examples of computer science work with big social media data: look at these papers and read the ones which interest you (skip the technical parts, and just go through general parts):

Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moo.
What is Twitter, a Social Network or a News Media? WWW 2010 conference.

Justin Cranshaw, Raz Schwartz, Jason I. Hong, Norman Sadeh.
The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City.
The 6th International AAAI Conference on Weblogs and Social Media. Dublin, Ireland,

June 2012.

Thomas Lombardi. The Classification of Style in Fine-Art Painting. 2005.

Michel, J. B., Shen, Y. K., Presser Aiden, A., Veres, A., Gray, M. K., The Google Books Team, Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., and Lieberman Aiden, E. Quantitative analysis of culture using millions of digitized books. *Science* **331,** 176 (2011).
If you are on campus, you will be able to read the whole article. If you off-campus, go to class Dropbox folder / TEXTS and you will find saved article there (without large size illustrations).

Recommended:

Chris Anderson. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. Wired 16.07.

Franko Moretti, Graphs, Maps, Trees: Abstract Models for Literary History - 1

selected papers which analyze twitter (dana boyd's list)

---

class 10:
final projects v1. review
examples of companies and research projects working with big social and cultural data
graduate students: big cultural and social data - research, possibilities and questions
undergraduate students: online portfolio basics / grad school applications

class plan:

undergrads: discuss portfolio preparation
grads: big data research research and questions

final project v.1 presentations
explain extra credit for CAPE evaluations

do department evaluations

Online portfolio preparation / graduate schools:

Elements of a Great Web Design Portfolio

Some of the popular free portfolio hosting sites:

Coroflot
Behance

Assembling a Competitive Application for Graduate Studies

Lists of best design schools (these are mostly for graphic design and/or industrial design, as opposed to media design or media art):

http://www.businessweek.com/interactive_reports/talenthunt.html

http://psd.tutsplus.com/articles/web/18-excellent-design-schools-from-around-the-world/

http://www.graphic-design-schools.org/schools/top-50-graphic-design-schools-and-colleges/

INNOVATE VISUALIZATIONS OF TEMPORAL PROCESSES:

Early time visualizations:

Time lines and visual histories

Visualizations of singular temporal streams:

One of the most famous visualizations of the last 200 years - Charles Joseph Minard's 1869 representation of Napoleon's 1812 Russian Compaign - offers other innovative solutions to show time:

Charles Joseph Minard - visualization of Napoleon's Russian Compaign

Last Clock - early influential project to show visual time at multiple resolutions

Visualizations of parallel temporal processes:

Here are a few examples of well-known innovative visualization techniques/projects to represent **multiple "event streams"** (multiple events which are taking place at the same time:)

History Flow

Last.fm listening history

Flight patterns

The Preservation of Favored Traces

Hans Rossling TED 2006 lecture (video on TED)

Ben Fry: energy use in a kitchen

(think also of Facebook Timeline design)

Visualizations of temporal links
(links between events in single or multiple temporal streams):

Map of Science

Citeology: Visualizing the Relationships between Research Publications

P.S. Also check bar chart demo programmed with Flare, currently the most popular software for programming visualizations

Visualizations of temporal processes in cultural artifacts:

Film Dialog Particles

Movie narrative chart

Cinemetrics project

Novel Views: visualizations of the novel Les Miserables by Victor Hugo

Lotr project

Evolution of Western Dance Music

Culture data visualizations by Santiago Ortiz

Artist Fellows: Visualizing Artists' Careers

Some software tools for visualizing temporal processes:

Check this bar chart demo programmed with Flare

Sparklines in Excel

Motion Chart in Google Spreadsheets

Timeline

Rickshaw

Resources: high quality visualization design:

I suggest that you read principles of information design summary (extracted from famous books on information visualization by Edward Tufte) - this will help you to come up with a better visualization design.

You can find lots of examples of high quality information visualizations on infosthetics.com.

Using a phrase "visualization design" in Google Image Search returns high-quality results (at least the first results page), so also try this for examples of good visualization design.

5-min Guide: Design principles for visualization

Over last 10 years, Information visualization has become of the key contemporary communication medium and also research method. But unless you went to design school, how can you create good looking designs such as these?

feltron.com

http://tulpinteractive.com/

http://www.pitchinteractive.com/beta/index.php

http://interactivethings.com/work/

http://tulpinspiration.tumblr.com/

(Note that I am not talking about visualization part itself - how to effectively translate data into visual representations, which visualization techniques to use when, etc. - but about graphic design part, i.e. how to "style" your visualizations.)

There are endless books and guides to graphic design, but they only work when you spend years practicing.

The purpose of this guide is to reduce modern design common sense to a few essential algorithms: mechanical principles which you should follow to arrive at a decent looking visualization design. Chances are you will not immediately come up with something as good as examples above right away, but at least you will start in the right direction.

(I came up with these "algorithms" while teaching classes in visualizations to my undergraduate students at UCSD during last for years. I noticed that every time we look

at their homework, same problems come and I give the same advice.. So I decided to put it all in one list.)

Here are **my six "algorithms" for designing good looking visualizations:**

1. Modern Design / infovis = systematic use of only a few options for each visual attribute:

Every information visualization design has a number of visual attributes. Common attributes include point style, colors, line widths, font family, font sizes, locations of text blocks, etc. The single most important design principle is: use only a few options for each attribute. For example, for your lines, use only maximum of two types of line width. For font size, similarly use use only two sizes: one for the title, one for the labels (for example). For color, use a palette of three or four colors; and so on.

2. Connect visual attributes with the semantics:

Connect these choices of attributes with the semantics of the visualization (i.e. what these visuals represent.) Each distinct visual option - a different line width, font size, a color, a position in a spatial grid, etc.. - has to indicate different type of content. Don't simply introduce more options to make your design "pretty."

3. Color:

Use of of the numerous web sites which contain professionally color palettes and color palette generators. Chose the one you like for your visualization and stick to it. Search for "color palette" or "color palette generator" to get to this sites.

4. Fonts:

Use no more than one typeface in your design, with only a few variations in size or style (two is better than three). Although this is just an example of algorithm 1, this is probably number 1 mistake beginners make - so I made into a separate principle.

5. Grid:

Another key principle of modern design is the use of grids. While they are less relevant when you create a single visualization, they becomes important than you start combining a number of visualizations together in a single design. Also use a grid when you are adding blocks of descriptive text to a visualization. (Again, this is a logical application of algorithm 1, its also one of the most common problems with the beginners.)

6. Take away what is not essential:

A famous Japanese product designer said: "When I design, I take away until there is nothing left to take out." When you design your graphs, ask yourself: do I really need to include labels in this graph? or grid lines? or axis lines ? or ticks? Simplify, simplify, and simplify again.

Examples of good web designs (for your project or whole blog / web site site):
Check these web sites for examples of modern minimal design aesthetics:

http://www.awwwards.com/websites/minimal/
http://www.minimalsites.com/

---

examples - use of visualization in museum and media collections interfaces:

Mace project
SFMOMA Artscope
Powerhouse museum australian dress register
Natural Science Museum of Barcelona heat map interface
Luna digital image collection software and collections

best musem / collection web sites:
Google Art Project
Walker Art Center
MONA

other interesting media collection visualizations and interfaces:
cinemetrics
0xdb.org
moviebarcode
Getty Images Moodstream

examples of museums putting their images and data on the web

examples of museum APIs / collection metadata:
Brooklyn museum
Cooper-Hewitt museum collection metadata
Powerhouse museum

aggregating museum data and images:
emuseum

museum analytics

Museum and the web conferences:
2012 (San Diego, April 11-14)
2011 conference program

UNIX for visualization design

introduction to Unix commands for humanities students (Scholars Lab, University of Virginia)

UNIX commands and techniques:
basic commands;
working with directories;

Generating a list of image filepaths with Unix commands (for ImageMontage and ImageSlice ImageJ macros)

some useful UNIX commands for working with big data sets:

find size of a directory (in kilobytes):
```
du -sk subdir
```

count number of lines in a file:
```
wc -l filename
```

copy all files from a directory which match some condition
(in this examples: all jpeg images)
```
find . -name *.jpg -exec cp {} /destination/dir \;
```

creating a histogram list of items in a file:
```
sort -n input.txt | uniq -c
```

remove selected columns from a data file:

1) Replace excel newline character with Unix newline character:

cat oldfile.txt | tr '\15' '\n' > newfile.txt

2) To remove 3rd and 5th column from a 8 column file:

cut -f1-2,4,6- oldfile.txt > newfile.txt

Software to analyzing and presenting online digital collections

documentcloud.org

viewshare.org

omega.org