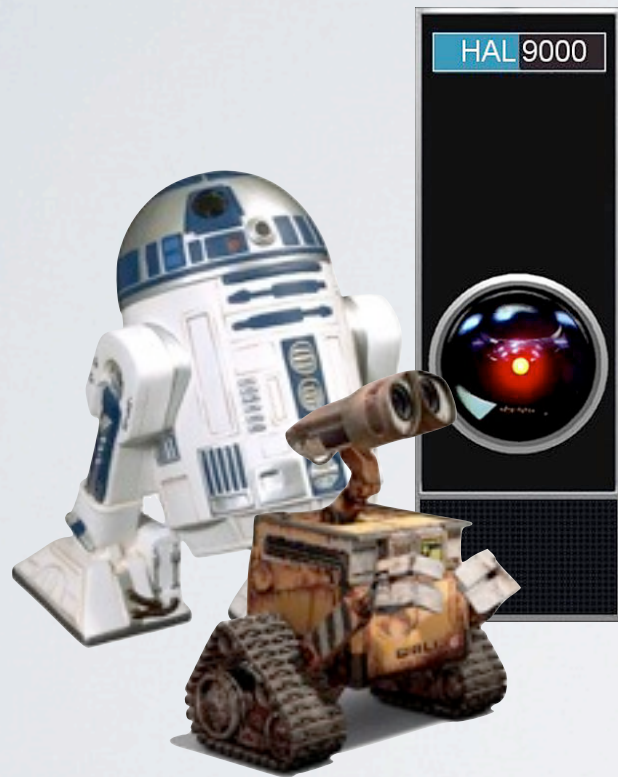


# LAST WEEK ON IO LAB



**Project 3** was due today at noon. If you haven't sent it to us and haven't already talked to us, come talk to us now.

Be ready to demo and discuss in class.

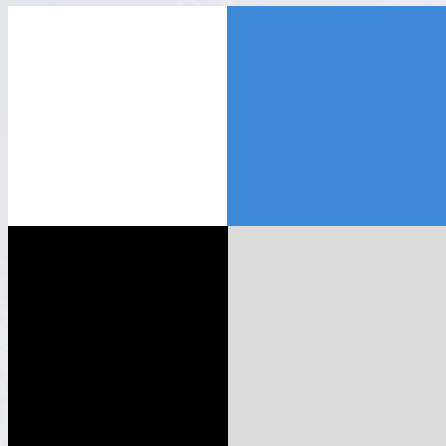


# INFORMATION ORGANIZATION LAB

Outline of the day.

# PROJECT 3 DEMONSTRATIONS





**delicious**



**flickr**

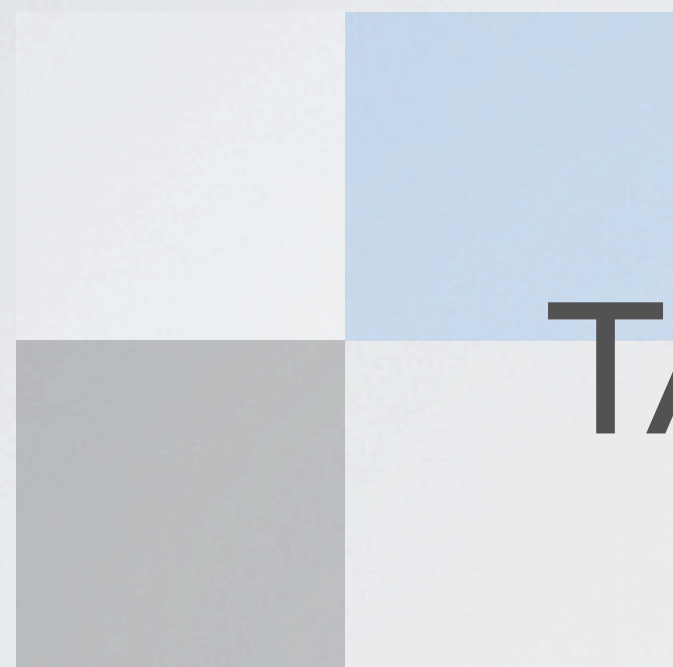
# PROJECT 4

Social & Distributed Classification

You must be thinking, “Oh, the work never ends.” Project 4 is intended to be a lighter project.

This project coincides with assignment 6 in i202. For those of you who took 202 last year (or haven’t looked at the assignment yet), this is the tagging assignment.

There are authoritative and canonical ways of organizing information. Vocabulary control, central authority.



# TAG USAGE



All projects are related to the same material as assignment 6, but think of this as the IO Lab companion to A6.

Typically we know that people tag things and the distribution: Zipf's Law, long tail, usage.



## Mr Eaves

Emigre has released a sans serif companion for Mrs Eaves, [Mr Eaves](#).

Mr Eaves was based on the proportions of Mrs Eaves, but Licko took some liberty with its design. One of the main concerns was to avoid creating a typeface that looked like it simply had its serifs cut off. And while it matches Mrs Eaves in weight, color, and armature, Mr Eaves stands as its own typeface with many unique characteristics.

Very handsome. I've always liked the attitude and flourishes of Emigre's typefaces. (via [quips](#))

By Jason Kottke • Oct 20, 2009 • [Post to Twitter](#) • [Emigre](#) [typography](#)

# SOCIAL VS. INDIVIDUAL

Compare the tags that an author applies to his or her own blog entries vs. those that other people assign those entries when saving links (on Delicious, Diigo, or other site). Note: this is only going to be an interesting comparison for very popular blogs.



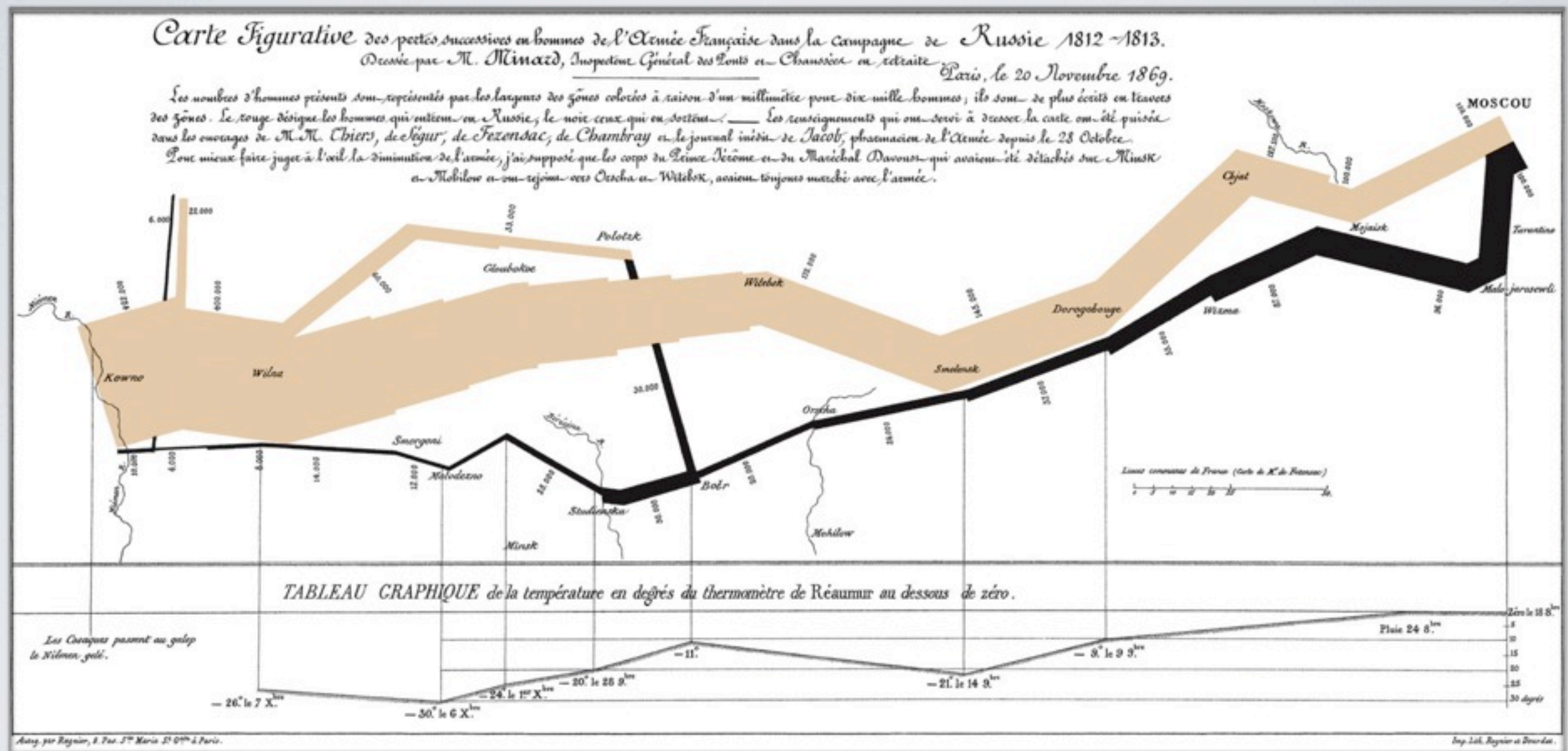
# CLASSIFICATION SCHEMES

If you remember what Yahoo tried to do with its online original directory, it used a system of people to categorize all the information on the Internet. How did that go? Maybe the real problem was that they didn't have enough people. Look at the categories developed from the group-up by Wikipedia: [http://en.wikipedia.org/wiki/Portal:Contents/Categorical\\_index](http://en.wikipedia.org/wiki/Portal:Contents/Categorical_index)

On Wikipedia there are thousands of user-created categories.

Demo starting at [http://en.wikipedia.org/wiki/Main\\_Page](http://en.wikipedia.org/wiki/Main_Page) and examine the categories in the top-right corner.





# VISUALIZATION

## Minard's Napoleon



# Raster

Pixels

Faster

Faster

Faster

**Faster**

# Vector

Formulas

Scalable

Smaller

Transformable

**Better**

Raster images are also called bitmaps.

It's said that "Raster is faster, but vector is better." People also say, "Raster is faster, but vector is correcter." But that sounds dumb.

# Canvas

Raster

2004 (Apple)

IE requires a plugin

Scripted bitmaps

# SVG

Vector

1999 (W3C)

IE requires a plugin

Declarative XML

- SVG = scalable vector graphics, like EPS in XML.
- You can provide support for canvas in IE without the end-user installing a plugin by using the excanvas library.
- An advantage of SVG is that each element is part of the DOM, and so you can manipulate individual elements, have events for them, etc.
- Canvas is part of HTML5.



# INTERFACE OPTIONS

Write raw XML or canvas code

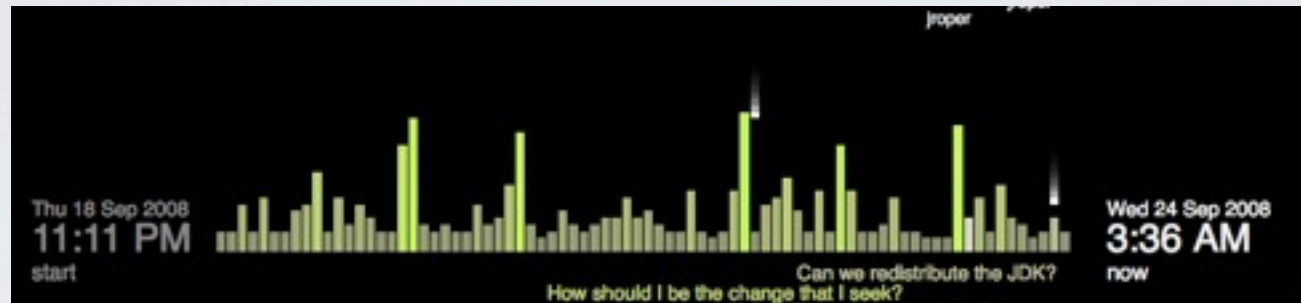


Use a visualization library



Use a chart and graph library

Raw canvas: [https://developer.mozilla.org/en/Drawing\\_Graphics\\_with\\_Canvas](https://developer.mozilla.org/en/Drawing_Graphics_with_Canvas)



# Processing.js

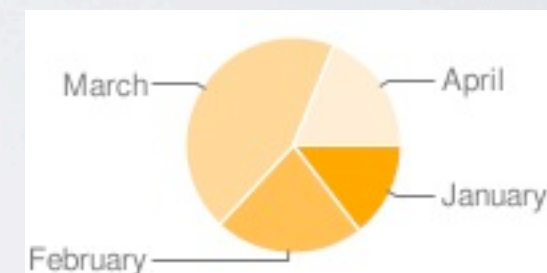
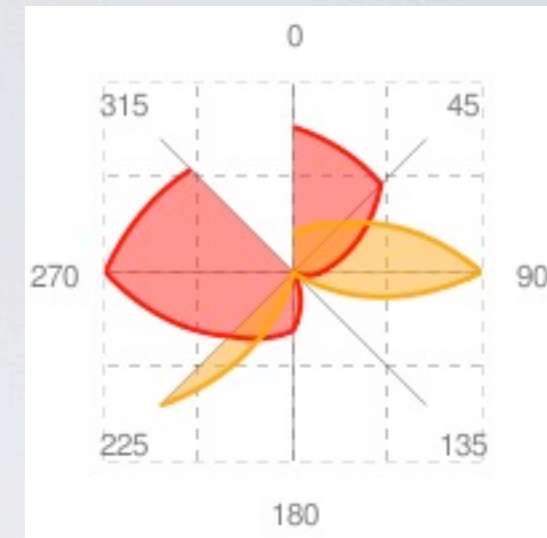
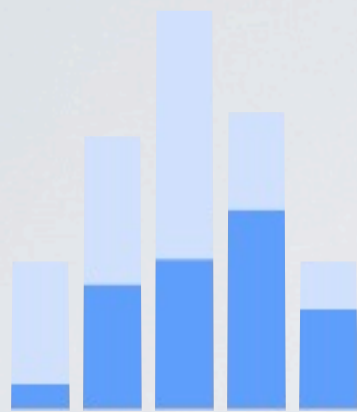
Processing and Processing.js are raster graphics libraries.  
Processing.js. Uses the HTML `<canvas>` element.  
These are general visualization toolkits (as opposed to specialized graphing/charting toolkits)





# Raphaël

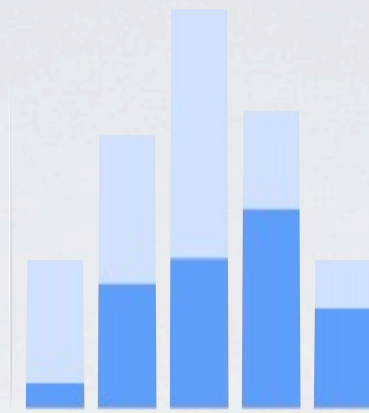
Raphael and gRaphael are vector graphics libraries. They use SVG elements.



# Google™ Chart API

Google—ever heard of it?





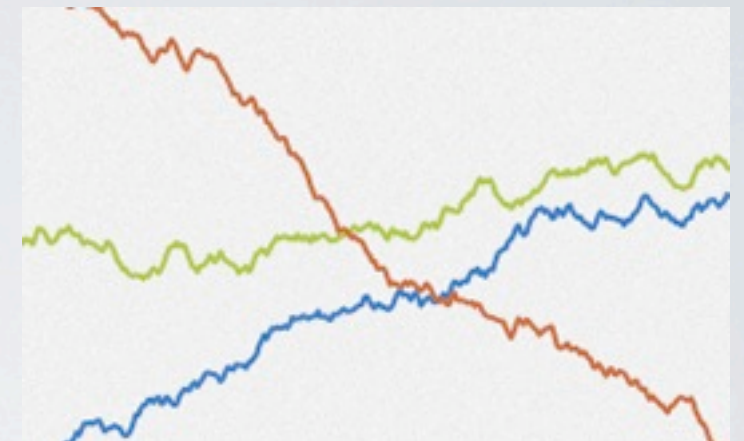
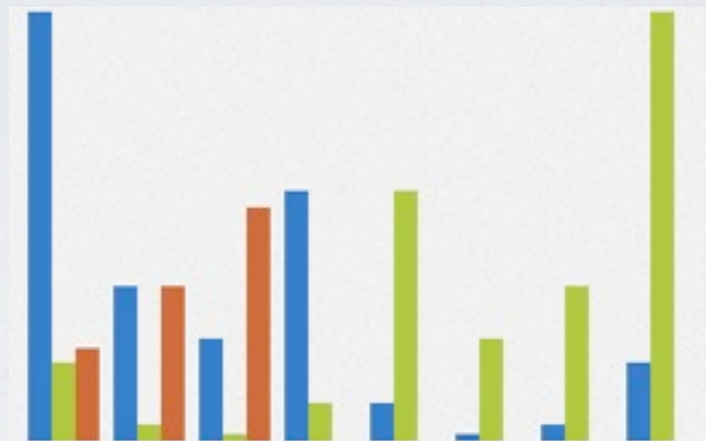
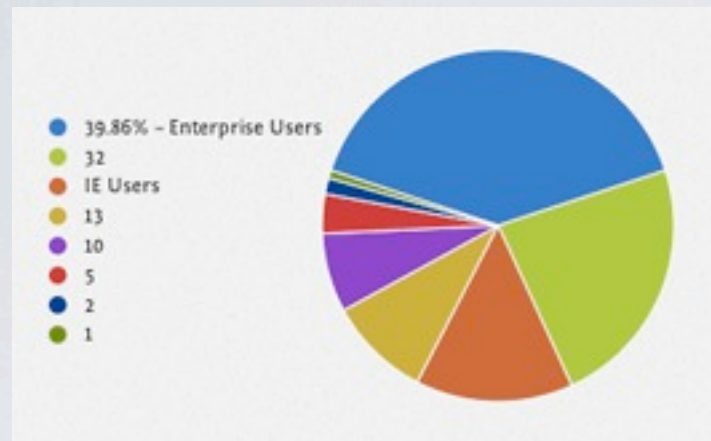
```

```

# Google Chart API

A simple bar chart

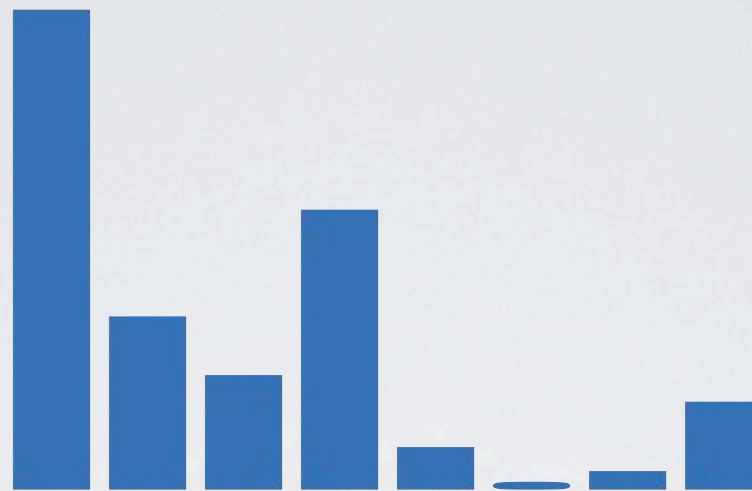
Doesn't use SVG or `<canvas>` -- it just returns an image!



gRaphaël

Raphael and gRaphael are vector graphics libraries. They use SVG.



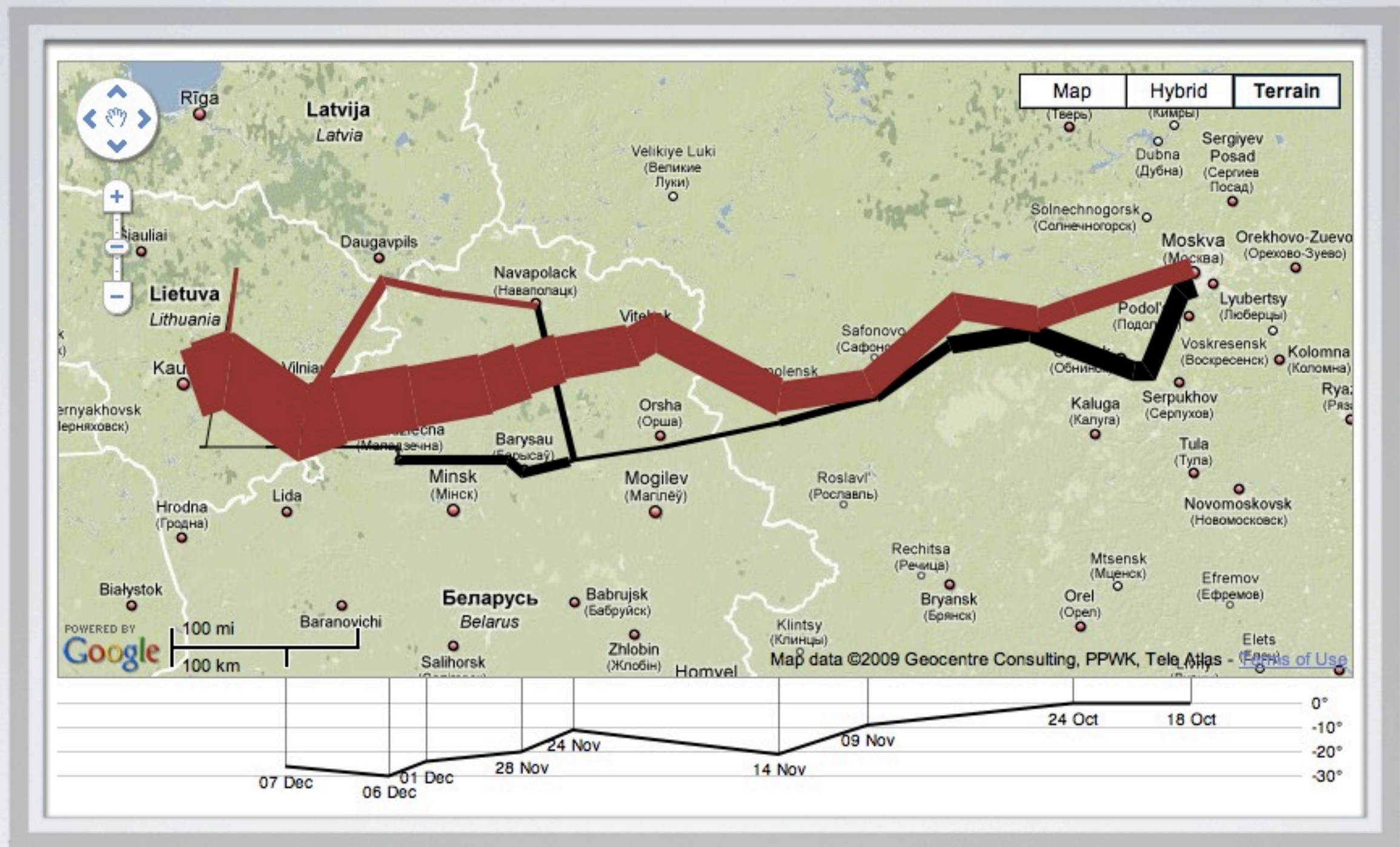


```
r.g.barchart(10, 10, 300, 220,  
  [[55, 20, 13, 32, 5, 1, 2, 10]],  
  0,  
  {type: "sharp"});
```

gRaphaël  
A simple bar chart

One line! Bare bones.



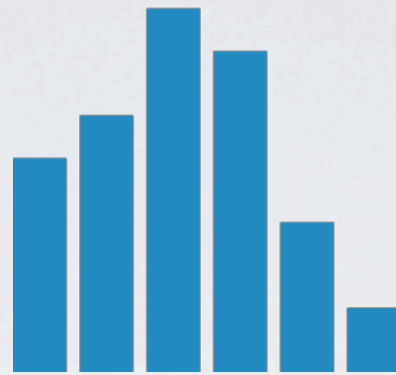


# PROTOVIS

## Minard's Napoleon

Protovis uses SVG. Developed by Jeff Heer (at Stanford, previously Berkeley).





```
vis.add(pv.Bar)
    .data([1, 1.2, 1.7, 1.5, .7, .3])
    .width(20)
    .height(function(d) d * 80)
    .bottom(0)
    .left(function() this.index * 25);
```

# PROTOVIS

A simple bar chart

It's chainable! Inline functions!

# GRAPHING DELICIOUS TAG FREQUENCY

This should help with 202 homework assignment 6. If you don't remember, assignment 6 is about using tags and tag frequency.

Demonstration in course resources on website at [http://courses.ischool.berkeley.edu/i290-4/f09/resources/tag\\_explorer](http://courses.ischool.berkeley.edu/i290-4/f09/resources/tag_explorer)



# CREATION PROCESS

- What am I trying to show?
- What data do I need? How do I get it?
- Data transformation
- Display

To make a basic visualization, you have to go through these steps.

Once you have the data, you need to get it into the format that the tool you're using wants. This usually requires creating an appropriate data structure, at least. It may also require some sampling transformation, like reducing the number of data points.

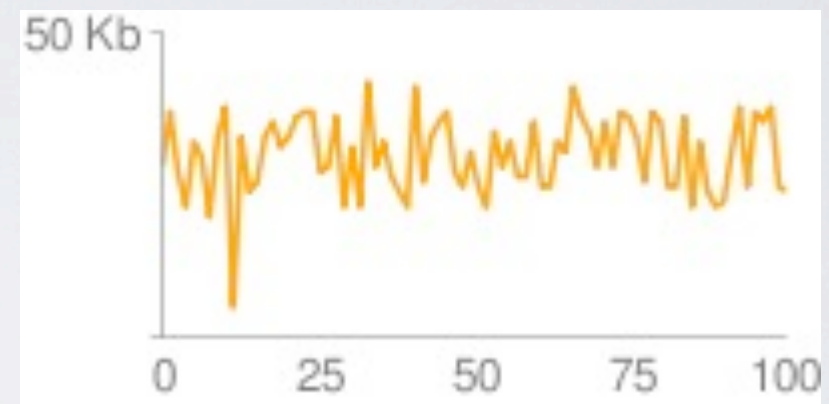
**Look at Delicious code.**

# GRANULARITY

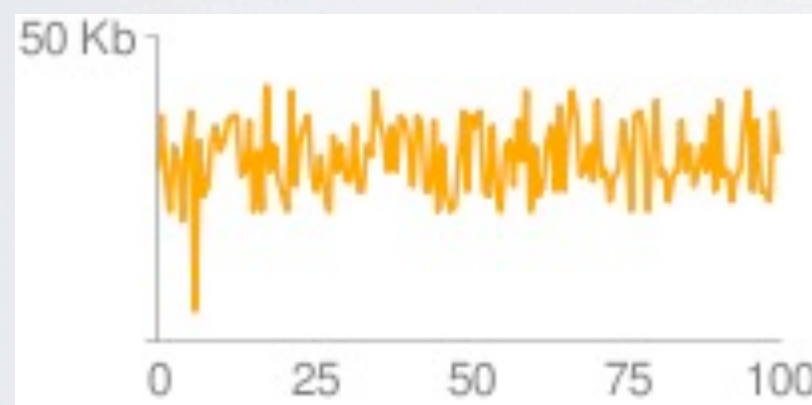
Don't store too many data points in too few pixels.



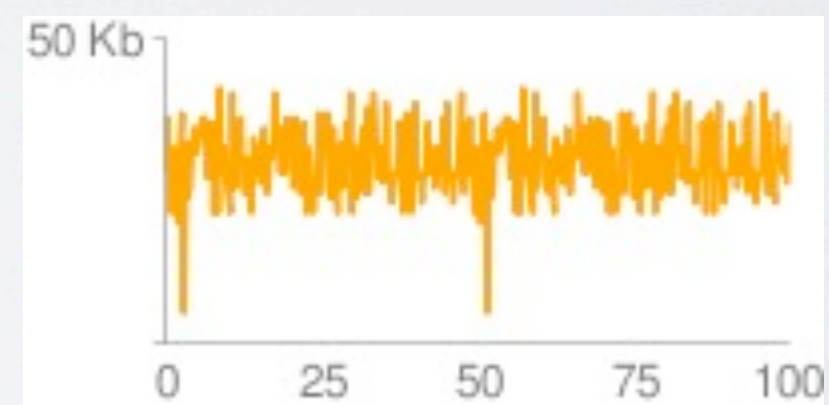
5 pixels per data point



2.5 pixels per data point



1.3 pixels per data point



less than 1 pixel per data point

Too few pixels is a concern even if you're using a vector format like SVG because ultimately the image is rendered at a certain screen size.

<http://code.google.com/apis/chart/formats.html#granularity>

This starts to get us into an issue that we'll discuss more next week.



# 202 EXAM GUIDE

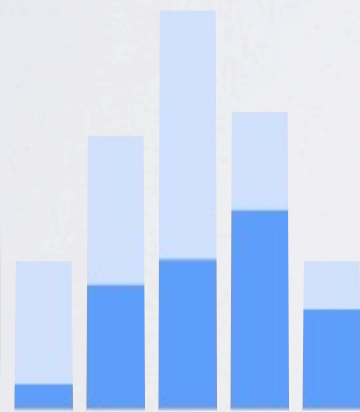
- There's some kind of tradeoff thing between organization and retrieval.
- There's a spectrum of documents from unstructured to highly structured."
- "These are hard problems, people."
- Answer questions using the phrase, "Based on an experiment I did in IO Lab..."

We know that a number of you are taking a midterm exam tomorrow. Good luck.

# FOR NEXT WEEK



**Find a group for project 4.**  
New people. They're out there.



**Check out a visualization tool.**  
Tools mentioned in today's lecture are linked on the course resource page.