Lecture 12:
Presenting our Results and Project Progress

COSC 526: Introduction to Data Mining
Spring 2020

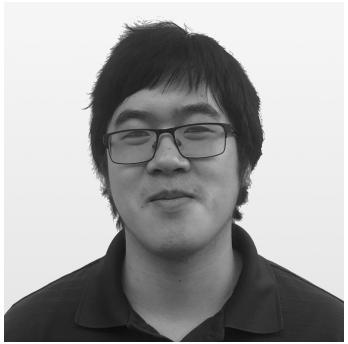
THE UNIVERSITY OF TENNESSEE KNOXVILLE

BIG ORANGE. BIG IDEAS.®

Instructor:

Michela Taufer

GRA:



Nigel Tan

Experts:

Leobardo Valera

Mike Wyatt

Project

Define your project (March 13)

- Which dataset will you be using? How are you obtaining the data? (We will provide you the appropriate data from NHANES or Medicaid, if you choose to use either of these datasets.)
- What is (are) the scientific question(s) that you want to answer? Be as specific as possible.
- What is your strategy to answer the question(s)? Define a set of steps that, if implemented with your code, will allow you to answer the question(s). Be as specific as possible.
- What is the tentative title of your project?

Create a new notebook with your solution (March 27)

- Write down the steps of your solution in distinct text cells; add one or multiple cells (as needed) to hold your code for each step. You can leave these software cells empty for the moment. Expand the text cells describing your solution.
- Add visualization cells that allow you to visualize results. You can leave these software cells empty for the moment.
- Add software to the code cells that upload data from source and pre-process data.
- Push your notebook into your GitHub repository as frequently as needed.

Finalize software and run tests within your notebook (April 3)

- Add the software that implements the method (or methods) to analyze your data.
- Add visualization cells that allow you to visualize results
- Push your notebook into your GitHub repository as frequently as needed.

Build a set of 15 slides that describe your work and get feedback (April 10)

- Build a set of ppt slides (use template provided) that summarize your work; use text slides to tell the story of your project and figures with the key results of your work.
- Make sure your slides include: motivation and problem definition, related work and background, your methodology (e.g., with flowcharts and code sections), your results, summary, and conclusions.

Create your poster and get feedback (April 17)

- Copy and paste your slides into the poster template.
- Shuffle as needed, extend and fill gaps, embellish fonts and text, enlarge text and figures to make them readable.
- **Submit your poster in GitHub (April 20)**
- **Present your poster at the virtual annual EECS COSC 562 poster session (April 24)**

How to prepare slides

Content

- Goal: present research results
- Think of the audience!
  - Knows foundation of e.g., bioinformatics
  - Does not know specific research
- Focus on concepts, not details!
- Point out advantages/disadvantages

Organization of a Talk

Title

Title of the Talk
Name
Affiliation
etc.

Splash

Fancy Picture
or
Fancy Quote

Summary

- important problem
- our solution is good
- techniques/methodology

most important

Overview

1. the problem
2. related work
3. notations
4. our solution
5. experiments
6. conclusion

Problem

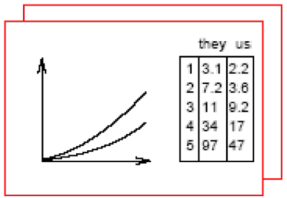
State−of−the−Art

Preliminaries

Solution

important

Tables



important

Conclusion

- summary
- outlook

most important

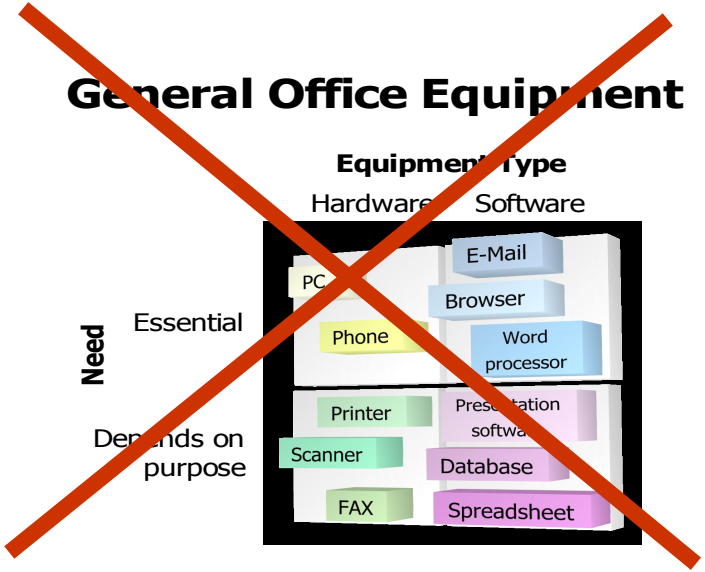
Giving the Talk

- Allow at least two minutes per slide
  - Not enough time? Cut it!
- Keep introduction and motivation short
- Use less than a minute on outline
- Use time for complex topics instead
  - Repeat if really important
  - Use examples

Slide Layout

- Keep it simple!
- Use Diagrams
- Try to keep it in one line
  - Many topics cannot be explained fully with a few bullets. In such situations, one is frequently tempted to write a full definition of the problem, usually in a very small font, which is read aloud while looking at the slide when giving the talk. Instead of this, a diagram together with a few key words would always be more suited to showing the context, which can then be elaborated on in more detail orally, in the talk.

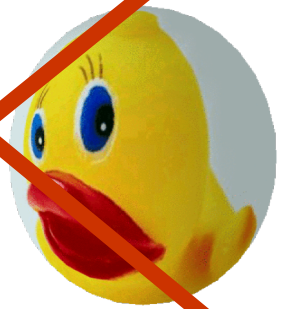
Diagrams



TOO FANCY

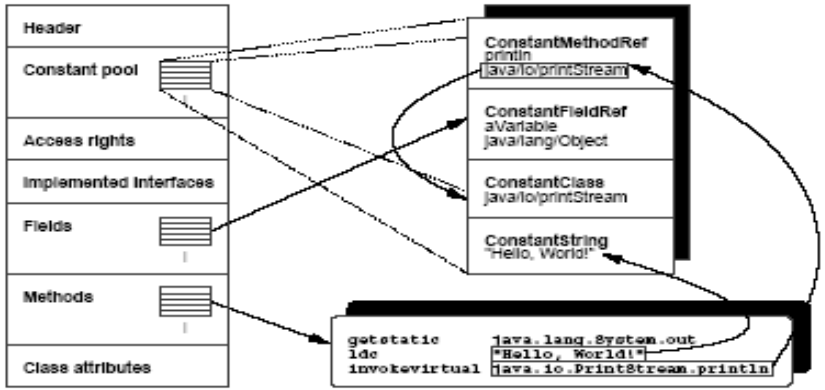
Ugly Duckling Syndrome

- Feathers
  - Stubby
  - Brown
- Told to leave town
- Low self esteem
- Does as is told



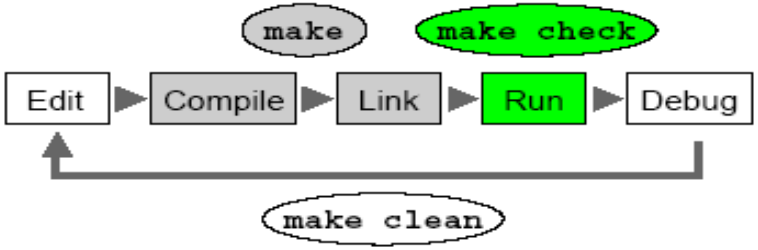
TOO ANNOYING

Diagrams

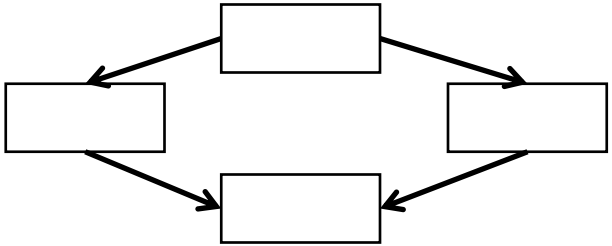


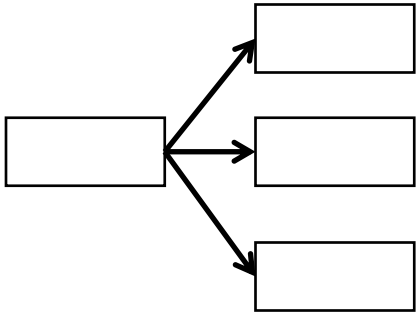

```
private static Method helloifyMethod(Method m) {
    /* Create instruction list to be
       inserted at method start.
     */
    String mesg = "Hello, World!";
    InstructionList patch  = new InstructionList();
    patch.append(new GETSTATIC(out));
    patch.append(new PUSH(cp, mesg));
    patch.append(new INVOKEVIRTUAL(println));
    MethodGen           mg  =
        new MethodGen(m, class_name, cp);
    InstructionList     il  =
        mg.getInstructionList();
    InstructionHandle[] ihs =
        il.getInstructionHandles();
    il.insert(ihs[0], patch);
    m = mg.getMethod();

    return m;
}
```

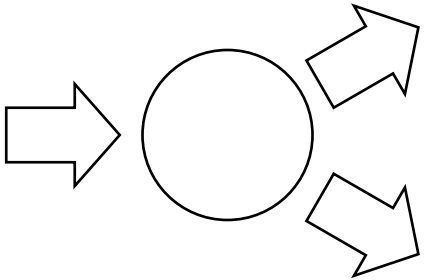
Think about the people in the back row!

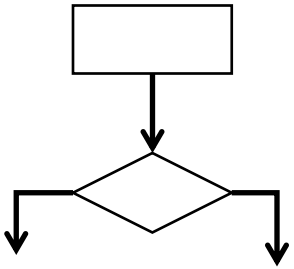


Structural Figures









Better than long text

Formulas

- Use only when needed
- Use descriptive variable names

$$velocity = \frac{distance}{time}$$

- Powerpoint equation editor is bad
  - Consider Latex + Acrobat Reader
  - Consider Texpoint

Colors

- Avoid too many colors
- LCD ≠ Projector
- Background colors:

  Yellow, orange, green, light blue

- Text colors:

  Blue, magenta, red, dark green

- Avoid distracting bitmap backgrounds

Project Progress

Project updates (I)

- Classification of distribution power system faults
  - Haoyuan
- Accelerating the Dimer Method with Machine Learning Techniques
  - Liubin

Project updates (II)

- Neutron Events Detection Using Clustering
  - Rebecca and SuAnn
- Using Different Clustering Methods on Single Cell RNA Sequencing Gene Expression Data to Determine Genes of High Importance For Human and Mouse Cells
  - Angelica

Project updates (III)

- Impact of soil moisture in wildfire simulations
  - Kae
- Freesound General-Purpose Audio Tagging
  - Pengxiag and Bohan
- Performance Comparison of Different MPIs using Hatchet
  - Ian

Project updates (IV)

- Authorship Identification Using N-Gram Feature Classification
  - Austin
- Medicaid Data Set from Delaware
  - Elizabeth and Aileen
- Stock Prediction
  - Abhijeet

Project updates (V)

- Detecting Trends in Twitter Health News
  - Nasib and Burcum
- Measurement of probable vulnerability of self-harm in different demographic using CDC survey data
  - Mohammad
- A Deeper Look into Scalable Methods for Creating Food Groups Using NHANES Dataset
  - Samuel

Question Queue

- Enter your questions here:

  https://tinyurl.com/s9xeqwh