

# Practical Experience for modeling work as statistical collaborators and consultants

A. John Bailer  
baileraj@MiamiOH.edu



MIAMI UNIVERSITY

# Outline

In my remarks, I will consider the general question of developing genuine statistical collaboration and consulting experiences in the curriculum

- Why is this important?
- Past patterns
- Present (& evolving) practice
- Reflection



# Why is Important? Context of remarks

## ASA Workgroup on Master's degrees in statistics

(<http://magazine.amstat.org/wp-content/uploads/2013an/masterworkgroup.pdf>)

Recommendations (based on survey of recent grads and employers):

1. Solid foundation in **statistical theory and methods**.
2. **Programming skills** critical and should be infused throughout the graduate student experience.
3. **Communication skills** critical and should be developed and practiced throughout graduate programs.



# Context of remarks

ASA Workgroup on Master's degrees in statistics  
(continued)

4. **Collaboration, teamwork, and leadership development** should be part of graduate education.
5. **Encounter non-routine, real problems** throughout their graduate education.
6. Internships, co-ops or other significant **immersive work experiences** should be integrated into graduate education.



# Context of remarks

## INGenIOuS (Investing in the Next Generation through Innovative and Outstanding Strategies) project (AMS,MAA,SIAM,ASA

<http://www.maa.org/programs/faculty-and-departments/ingenious> )

1. Bridge gaps between business, industry, and government (BIG) and academia.
2. Improve students' preparation for non-academic careers - *better preparation will increase the number of graduates who enter the workforce well equipped with skills and expertise in mathematics and statistics. Change is needed both in curricula and in some faculty members' perceptions of BIG careers for their students.*
3. Increase public awareness of the role of mathematics and statistics in both STEM and non-STEM careers.



# Context of remarks

## INGenIOuS (continued)

4. Diversify incentives, rewards, and methods of recognition in academia - *A well-balanced mathematical sciences program offering a bachelor's degree or above should include faculty with a variety of interests: **discovery research** (in pure and applied mathematics and statistics and mathematics education); **work in applied, collaborative, and interdisciplinary areas**; and **teaching and preparation for careers both within and outside of academia.***
5. Develop alternative curricular pathways.
6. Build and sustain professional communities.



## Past Patterns for data practicum classes

Initially, common features of early versions of data practicum classes included:

- Problems described and motivated by the instructor using artificially clean preprocessed data
- Labs were previously analyzed and a particular solution is likely
- Students alternated presenting by all students submitted reports of each analysis (often initial + final report)
- Stat instructor provided all feedback to oral/written reports



# Past Patterns for data practicum classes

A company was thought to be polluting a local lake by discharging its manufacturing waste into the lake without pre-treatment. To investigate whether the lake was polluted, the EPA took five samples from the lake receiving the discharge (Lake #2) and five samples from a nearby unpolluted lake (Lake #1). Strontium measurements were recorded for each of the samples.

Data:

Lake #1: 27.2 29.1 33.2 31.4 32.8

Lake #2: 37.4 35.0 41.2 40.6 36.2

Goal:

Determine whether the strontium concentrations are different for the two lakes.

Requirements:

Provide both graphical and numerical summaries as part of your analysis. All reports must be typed. Line printer plots are NOT acceptable.



# Past Patterns for data practicum classes

## Strengths of historical structure:

1. Labs could be designed to span a breadth of statistical methods
2. Class was usually small (3-8) and students had lots of chances to present.
3. Opportunity to present ideas not formally covered in other classes.
4. Relatively homogeneous student population (teams made up of students with similar majors)



# Past Patterns for data practicum classes

## Weaknesses of historical structure:

1. Problems were already well formulated by a statistician – no need to translate problem from a client
2. Data were preprocessed and relatively easy to mold into an analysis data set
3. Relatively homogeneous student population



# What do employers want?

Quick review of postings on 'indeed.com.my' for 'statistician'

Data Mining Specialist in Kuala Lumpur position

The Successful Applicant will have ...

- At least 5 years of experience in financial services industry
- Implementation experience in data mining and data processing methods
- Advanced knowledge of SQL and relational databases; SAS experience pref.
- Degree in IT, Quant Methods, Econometrics, Mathematics, Comp Physics { *statistics implied by quant methods?* }
- Good communication – both verbal and written in English, able to communicate across internal and external stakeholders
- Competent, committed and matured professional



# Present (& evolving) practice

Current context –

Moved undergraduate data practicum and graduate data practicum courses to client focused

Added data visualization class with multidisciplinary teams working on projects



# Present (& evolving) practice

## STA 475 Data Analysis Practicum (3) MPC

The use of statistical data analysis to solve a variety of projects. Emphasis on integrating a broad spectrum of statistical methodology, presentation of results both oral and written, use of statistical computing packages to analyze and display data, and an introduction to the statistical literature. A term project involving student teams combines elements of all of the above. CAS-QL. Prerequisite: STA 463/563 or 363; or ISA 291.

## STA 660 Practicum in Data Analysis (3)

Supervised practice in consulting and statistical data analysis including use of computer programs. Maximum of six hours may be applied toward a degree in mathematics or statistics. Offered credit/no-credit basis only.

## STA 404/504 Advanced Data Visualization (3)

Communicating clearly, efficiently, and in a visually compelling manner using data displays. Identifying appropriate displays based on various data characteristics/complexity, audiences, and goals. Using software to produce data displays. Integrating narratives and data displays. Critiquing visualizations based on design principles, statistical characteristics, and narrative quality. CAS-QL. Prerequisite: at least one of the following: STA 261, 301, 368, 671; IMS 261; ISA 205; or by permission of instructor. Cross-listed with IMS/JRN.



# Present (& evolving) practice

## Desire:

1. Direct engagement in wrestling with client-defined tasks
2. writing outcomes
3. group work
4. service learning



# Present (& evolving) practice

## Challenges and implementation

1. Getting Clients?
2. Projects
3. Reflection

# Getting Clients?

Need to actively recruit clients and screen projects  
Targeted email ...

SUBJ: An invitation to propose projects for data analysis capstone /  
practicum class

Greetings,

Have you or your office collected data that you haven't had the chance to analyze? Are you planning for future studies and would like some assistance determining how many observations you might need? Do you like working with motivated students? If you can answer "yes" to any/all of these questions, then I invite you to put my students to work ...



# Getting Clients? (continued)

If you have a project(s) where statistical assistance might be valued, then let me know. Please send me a short description of the project including:  
(on the email subject, please use the convention - SUBJECT: STA 475 project: your name - project title)

1. Short descriptive project title
2. Goal of the analysis (e.g. design phase - project planning; data analysis, etc.)
3. Data to be analyzed (e.g. Excel data sheets; still to be collected; ...)
4. Type of statistical analysis anticipated (e.g. logistic regression, anova models, etc.)
5. Timeline for analysis (when are results needed)



# Getting Clients? (continued)

Observations:

Once you do this once, clients will return in future.

Current repeat clients in my data practicum class:

- Gerontologists

- Exercise Physiologist

Current repeat clients in data visualization class

- Local paper – Cincinnati Business Courier

- Research center – Scripps Gerontology Center



# Projects

I am the first client (probably the worst they will have)

Project: compare dissolved oxygen-depth relationship between two lakes

Very general guidance on first draft report

Extensive commenting on first draft to be addressed with revision (at least one revision)



# Projects (continued)

Ideas to convey early ...

1. Revision and critical reading of reports key skill and learning outcome
2. Better graphical displays lead to easier writing and communicating with clients
3. Reporting effect estimates often richer than exclusively reporting the results of hypothesis testing { *indicates what students are taking away from our classes* }
4. Writing a structured report is a skill



# Projects (continued)

Later projects – teams; charter-traditional schools; CELTUA; OMA; development)

- Project life course
- Client meeting and discussion with class
- Team work
- Interim reporting of analyses
- Draft report produced (class wiki for other teams to comment)
- Client presentation of final analysis
- Final report and presentation sent to client



## Projects – Example OPTAB

Task: conduct an occupancy study of the various parking areas in town so that the Oxford Parking & Transportation Advisory Board (OPTAB) could make data-based decisions regarding meter rates, hours, fines, etc.

Client: board which represents nearly every constituency in town: City Council, Miami University, Chamber of Commerce, Oxford Landlords, Mile Square Residents, Year Round Residents, ASG, Talawanda Schools and McCullough-Hyde. Ex-officio members include Chief of Police and the City Manager



# Projects – Example OPTAB

## Issues:

When a block face, i.e., the collection of parking spaces on a street for a city block, exceeds 85% occupancy it appears full.

If the remaining 15% of block face spaces are vacant, then this would lead to a loss of earnings (if these are metered spaces).

City parking garage has low occupancy.



## Projects – Example OPTAB

Conduct an analysis of a sample of the 790 metered spots in Oxford as well as the parking garage.

Goal: investigate different rates based on location to spread occupancy from “hot spots” to outer locations as well as into the unoccupied city garage. In addition, ‘meter feeding’ could be removing spaces from circulation for extended periods of time.

Question: extending meter hours past the current 6 p.m. deadline which could force many vehicles to park away from the high and main intersection as well as produce additional income for the City of Oxford.



# Projects – Example OPTAB

## Solution:

Developed a data collection plan that was implemented in March of 2010.

Four days (Monday, Thursday, Friday and Saturday) were sampled hourly (11 a.m.-3 p.m.; 4 p.m.-8 p.m.) - restricted attentions to 27 block faces and the parking garage for a total of 377 parking spaces.

## Logistics of the data collection:

- order that spaces were checked;

- data collection sheets;

- decisions about what to record such as license numbers)

- appropriate documentation and emergency numbers.



# Projects – Example OPTAB

## Solution:

Developed a data collection plan that was implemented in March of 2010.

Four days (Monday, Thursday, Friday and Saturday) were sampled hourly (11 a.m.-3 p.m.; 4 p.m.-8 p.m.) - restricted attentions to 27 block faces and the parking garage for a total of 377 parking spaces.

## Logistics of the data collection:

- order that spaces were checked;

- data collection sheets;

- decisions about what to record such as license numbers)

- appropriate documentation and emergency numbers.



# Projects – Example OPTAB

For each metered spot, occupancy was recorded and whether the meter was in violation. Summaries of the collected data included:

- Occupancy/Availability rate per time, day and block on street parking

- Occupancy /Availability rate per time, day in garage

- Violation rate per time, day and block on street parking

- Violation rate per time, day in garage

- Meter feeding: less than 2hrs, occupied over 3hrs including location

Results reported to the OPTAB clients included:

- Satellite Map with highlighted lines

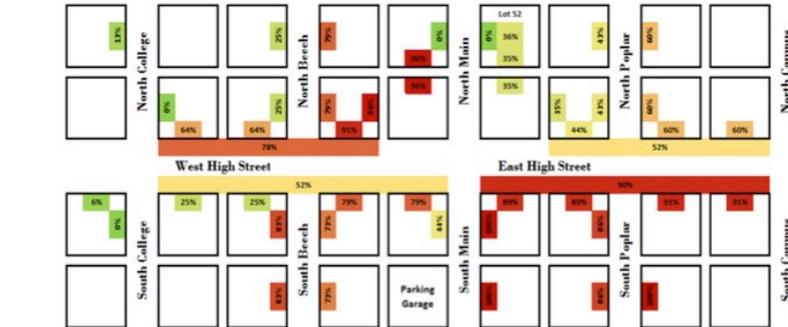
- Heat map



# Projects – Example OPTAB

Example of Heat Map of occupancy for one sampling time

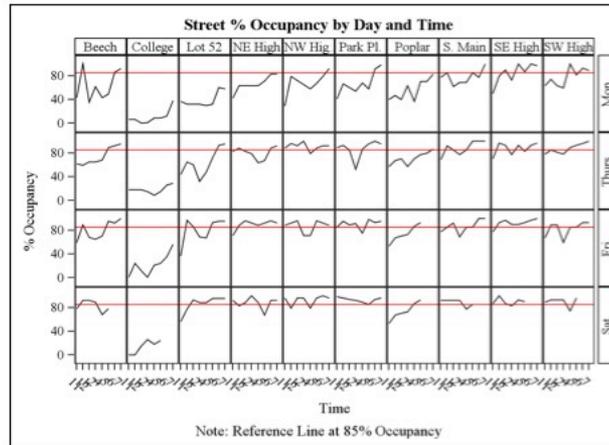
Monday – 1pm



# Projects – Example OPTAB

Occupancy of block face over time on different days

## Daily and Weekly Trends



# Projects – Example OPTAB

## Impact

Students presented the results of this data collection and analysis effort to the OPTAB. The board was very impressed with the depth, quality and insight provided by this project. The board mentioned that these data would be relevant for setting parking meter rates that might include differential rates for underutilized spaces.

Finally, the OPTAB members commented that this level of work may have exceeded the value of a previous report that cost over \$20000 to conduct. At this point, the students were left speechless.



# Projects – Example OPTAB

## Follow-up

Experience of working with a local government organization and contributing to information needed to support decision making

Students had to coordinate an extensive data collection effort, process the data into an analysis data set, construct displays and summaries that gave insight and develop a presentation and report that captured all of this work.

Comment about the cost of previous parking studies led to interesting class exercise of developing a cost estimate of how much billable work was reflected in the analysis, and what they would charge if they did this as a consulting company



# Reflection

## Assessments:

### Learning outcomes (BS STA degree)

1. Students shall be able to analyze and interpret data critically using statistical models and programming skills *{ also part of practicum }*
2. Students shall demonstrate understanding of the mathematical basis and theoretical foundations of statistics
3. Students shall be able to effectively communicate, both orally and in written form, results of statistical analyses to both the expert and layperson *{ data practicum class critical data for evaluating this LO }*



# Reflection (continued)

## Reflection - STA 475 Data Practicum Portfolio and Self evaluation

“Each **Capstone** emphasizes **sharing of ideas, synthesis, and critical, informed reflection** as significant precursors to action, and each **includes student initiative in defining and investigating problems or projects**” (<http://www.units.muohio.edu/led/Capstone>). Your final exam is a paper reflecting on the following questions derived from the reflections associated with service learning courses.

*What?*

What **projects** did you work on this semester?

What was your **role** on project teams? Note that this may change on different projects.

What did you **observe**?

What did you **like/dislike** about working on these projects?

What **missing knowledge or skills** would have made you a more effective contributor to the work on these teams?



# Reflection (continued)

## *So What?*

What stuck out about the experience in this capstone?

What was the **best/worst** thing that happened?

What have you **learned about yourself**?

How does this **experience compare** to others you've had?

## *Now What?*

What have you **learned** about **working** as a statistical **collaborator**/consultant?

How did this experience **challenge** you?

Do you believe that this course **prepared you for future collaborative work**? If so, how? If not, why?

## *Summary*

What **grade** would you assign to your efforts this semester? Why?

Describe the work or contribution of particular **individuals** to project teams that you thought was **noteworthy**.



## Reflection (continued)

*Reported from students ... during interviews with potential employers*

*Employers expect graduate will have the stat and math background but look for the experience with client (often messy) problems, where they worked in teams and had to produce written and oral reports for the clients.*

Experiences such as described above provide this and I encourage you to develop (as academics) and to support (as potential clients)



Thank you

