Data Wranglers:
Leveraging Graduate Student Expertise for Data Support

Mitch Winterman MLIS
University of Nevada, Reno
What is your Library doing to support Data?

- Help Finding Data?
- Research Data Management?
- Data Visualizations?
- Something Else

What is your staffing model for this service?
• Literature
  – Most Research Data Services research focuses on RDM (Cox, Kennan, Lyon, & Pinfield, 2017)
  – RDM research focus places emphasis on experienced scholars already proficient in Data (Beauchamp & Murray, 2016)
  – Teaching Data Skills and statistical software packages (skills novices require) has not traditionally been in the domain of librarians. Usually learned through academic courses, self directed study, or workshops like the Carpentries (Dekker & Lackie, 2016)
  – Five areas of Data-lifecycle: project startup, data collection, describe phase, data analysis, & data Sharing. Library services: RDM services, DMP services, metadata consultations, data archiving, institutional repository provision, and data sharing support. (Yu, 2017)
• Literature (cont.)
  - Most libraries are not operating within the data analysis lifecycle because they have not been able to create an alignment of the staff, resources, and space necessary for such an endeavor (Guss, 2016)
  - Libraries that are offering data analysis support use professional staff either by retraining existing staff or recruiting (Cox, et al., 2017)
  - Case Studies
    • Penn State – Data Learning Center & Research Data Enclave – Professionally staffed
    • Duke University – the EDGE – Professionally staffed
    • University of Colorado, Denver – Creative Technology Commons – Professionally staffed
    • New York University – 10 professional staff (two managers, 3 quantitative data/statistics specialists, one qualitative/survey specialist, two GIS specialists, one GIS librarian, one Data Librarian, and six to eight graduate student consultants)
    • University of North Carolina, Chapel Hill – Research Hub – staffed by specialized librarians, technology specialists
    • University of Virginia – Originally staffed by a graduate student, but that model was unsustainable and is now staffed by multiple professionals
• Background
  – Spring 2018 - Brainstorming how to support Data better
    • Research Data Management support
    • Computer lab space
  – Tech Wranglers
    • Makerspace Student workers
    • Help with 3D printing, 3D modeling, laser-cutting, hand tools, and more
  – Proposal
    • 5 W’s
I think this would be an incredibly helpful resource for our undergraduates and graduate students! Especially SPSS and R. I definitely think students would use it.

- Dr. Marin Pilloud, Anthropology

I think that is a great idea! And, my students would definitely take advantage. I have frequently heard them say that they are grateful for the computers with SPSS installed, but wish that those on duty could answer questions about the software. Please keep me informed of what you decide and I will pass it along each semester to students in my EDRS640 class.

- Dr. Cleborne Maddux, Education

I think this is a good idea - grad students can brush up on their SPSS skills and deepen their understanding by helping students problem solve. However, I have noticed a range in competency when it comes to statistics at the grad level. Most of the undergrads in my PSY 240 class have never used SPSS and do not feel confident using statistics, so having a space to consolidate what they learn in class would be great!"

- Dr. Sarah Haigh, Psychology

Students would take advantage of this service if it were available! I think it’s a great idea. Can you let me know if this will be implemented this semester? We have 6 students in an advanced machine learning (in R) class right now and being able to guide them to the library for extra assistance/questions would be a huge time saver for me! I can also let their instructor know so he can announce it to the class.

- Dr. Kyra Stull, Anthropology
I think this would be very useful. Students often spend a lot of time and get frustrated when encountering coding issues. Often they spend too much time trying to figure out a coding/syntax issue rather than focusing on the content and interpretation of what they are doing. Frequently I find that it is a simple solution, but they spend a lot of time due to their unfamiliarity with the software. I also sent this to a few other faculty that teach classes using stat software.

- Mark W. Nichols, Economics

I think we chatted about the possibility of such a thing in Spring. I do think that there are a number of students who would take advantage of a service like this for (a) thesis projects, (b) some research projects (both student or faculty led), and (c) students who are struggling with technical aspects of homework in stats classes using these programs.

To give this some context, every year or so I have some desperate grad student who cannot figure out their own thesis data. And I used to be part of a stats consulting service that the Center for Research Design and Analysis used to offer, where there were always people in need of support. Assuming that I only see a fraction of the people (and typically only the tougher cases out of certain programs), I speculate there must be quite a few folks out there. I also think that Yueran Yang (though only entering her 3rd year as assistant professor) has had similar experiences; heck, even I have hit her up for some help with R code when she was brand new!

There’s also the fact that for our first-year statistics course PSY 706 (taught every Fall), we have had one 10-hour TA to support students with R. Again, this course is bigger than it should be (25 students), and the TA’s lab hours will not be enough for everybody. And that’s only one course, not even considering the various other places on campus were R is used in undergraduate and graduate education (geography and biology come to mind). And SPSS is of course still popular elsewhere on campus (e.g., education).

In short: YES, I think this will be a popular service once word gets out. I do think our own students might draw on this kind of a support primarily during their first year, when many are learning R, even though we as a program try to provide some technical support. Once they are a bit more advanced, I see our students much more in the role of providers of statistical support, simply because our students get more statistical training & experience than other programs around us. There are also programs with little or no stats support (e.g., engineering education), which would benefit from this service, I think (even when these folks also work with qualitative methods).

- Dr. Markus Kemmelmeier, Social Psychology
I am a professor for ECON 441 (Intro to Econometrics), and Mark Nichols forwarded me an e-mail of yours that said the library was considering offering help to students with statistical packages like Stata. I conducted a brief survey in my class to gauge potential demand, and around 25 students were interested. Is there any news regarding the initiative? Could you please keep me updated?

- Dr. Peter Toth, Economics

Based on conversations with my students, yes I think the help would be welcome. It would also take the pressure off of Kyra and Marin, both of whom are in great demand for R help.

- Dr. Geoffrey Smith, Anthropology

I believe that the service will be helpful for graduate students. I know that many graduate students seek help when they conduct their studies. I suggest that the service is provided by graduate students who not only know how to use SPSS but have advanced training in educational statistics.

- Dr. Li-Ting Chen, Education

I teach Economics 741 to 16 students, which uses both Stata and a bit of R. I think that this would be a great idea and very useful resource for my students.

- Dr. Todd Sorenson, Economics

I think if you have the budgets to do so, it would be good to have the assistance. You need to find someone who is really good at the basic statistics methods and SPSS. Just wondering whether this would be a student worker or a GA position.

- Dr. Leping Liu, Education
• What the heck is a Data Wrangler?
  – Graduate Student Worker
  – Already has expertise regarding both statistics
    AND at least one statistical software
  – Experience/skills in teaching
  – Specialized skills = more $$
  – Work 4-10 hours/week
    • 4 hours of walk-in hour support
    • Up to six hours of appointments
• Wrangler #1
  – Hired November 2018
  – PHD candidate in Political Science
  – Excellent Employee
  – Expertise in R and Stata
  – Interest in future career in academia
• Wrangler #2
  – Hired March 2019
  – Masters student in Information Systems
  – Interest in future career in Data Science
  – Discipline issues (missed appointments, tardiness, lack of productivity, possible lack of expertise)
  – Terminated in August 2019
When not performing consultations:

• Internal workshops
• External workshops (GSA, Post-Docs lunch and learn)
• LibGuides
• Working with internal library data
Assessment:

• What were the perceptions of the patrons who received consultations? Were they satisfied and did the appointments meet their expectations?

• How did the student workers feel about their work? Were they confident that they were able to help users?

• Was the service being utilized to its full potential? How were patrons using the service? What was more impactful, walk-in assistance or consultations booked in advance?

• How can the program be improved? What changes, particularly in the area of training, make meaningful impact?
Wrangler Survey - Findings

• Very confident they had the knowledge, skills, and expertise to assist patrons

• Suggested improvements:
  – More weekend workshops to train more people and decrease one-on-one demand
  – More hours available for consultations
  – Learn more skills
Wrangler Survey – Findings (cont.)

- Felt the role would help with future academic endeavors and career aspirations
- Wranglers felt users probably had satisfaction rates between 75-100%
- Patron Dissatisfaction may occur if:
  - Limited consultation length or frequency
  - Lack of expertise on specific software
  - Cognitive dissonance between service offered and user expectations
Patron Survey – Findings

• (Survey was sent to both tech and data wranglers – Results reflect Data Wranglers unless otherwise noted)

• At the time the survey was sent (Summer 2019) there were 27 unique patrons accounting for 58 booked consultations

• 12 users responded – 45% response rate (tech wrangler response rate: 31/360 – 8.6%)
• Patron Survey – Findings (cont.)
  – Patron Type

![Chart showing patron types](chart.png)
Patron Survey – Findings (cont.)
– Marketing

Which services

- Tech Wrangler (DeLaMare Science & Engineering Library: Makerspace)
- Data Wrangler (Mathewson-ICT Knowledge Center: @One)
- GIS Tutor (DeLaMare Science & Engineering Library: Basement)
- Other

Legend:
- Librarian or other library staff
- Library website
- Friend/colleague
- Instructor/Professor/Advisor
- Other
Patron Survey – Findings (cont.)
– Project type

Which services

- Tech Wrangler (DeLaMare Science & Engineering Library: Makerspace)
- Data Wrangler (Mathewson–IGT Knowledge Center: @One)
- GIS Tutor (DeLaMare Science & Engineering Library: Basement)
- Other

Academic (Class or Research-related) | Personal | Professional | Other
Patron Survey – Findings (cont.)
– Number of Repeat Consultations

Which services

- Tech Wrangler (DeLaMare Science & Engineering Library: Makerspace)
- Data Wrangler (Mathewson-ICT Knowledge Center: @One)
- GIS Tutor (DeLaMare Science & Engineering Library: Basement)
- Other

![Bar chart showing the number of repeat consultations for different services.]

- 1
- 2–4
- 5–7
- 8+
- Other
Patron Survey – Findings (cont.)
- Needs and Satisfaction

![Graph showing needs met and satisfaction with expertise and knowledge.]

Number of Responses

Yes | No | Somewhat

Needs Met?

Were you satisfied with the level of expertise and knowledge your Wrangler had?

Number of Responses

Yes | No | Somewhat
Patron Survey – Findings (cont.)
– Recommend to Friend/Colleague

Would you recommend a Wrangler Consultation to a friend or colleague?

Number of Responses

Yes | No
Patron Survey – Findings (cont.)
– Apply What was learned

Are you likely to be able to apply what you learned during your consultation in the future?

Number of Responses

- Yes
- No
- Unsure
Survey – Issues

– Only 2 Data Wranglers – limited candor?
– User survey only hit those that pre-booked
  • Walk-in users could not be surveyed
– Who responds to surveys?
  • 45% response rate
  • Typically those who love or hate something
Wrangler Questionnaire Form

• Online form Wranglers fill out after every consultation (booked and walk-in)
• Supposed to act as a master file to record all meaningful patron interactions
• Data Wranglers used this form to create 45 entries
Wrangler Questionnaire Form (cont.)

UNR Affiliation:

- Undergrad
- Masters Student
- PHD student
- Faculty
- Staff

Values
Wrangler Questionnaire Form (cont.)

How did they learn about this service?

- Library website
- Professor recommendation
- Referred by service desk
- Librarian/Library
- Walk-up
- Don't Know

How did they learn about this service?
Wrangler Questionnaire Form (cont.)
Wrangler Questionnaire Form (cont.)

What software/equipment did you work with?

- R: 37.50 %
- SPSS: 20.83 %
- STATA: 27.08 %
- Python: 6.25 %
- Excel: 8.33 %
Wrangler Questionnaire Form – Issues

• Data did not match real number of consultations
  – There were a higher number of bookings (58) than entries (45)
  – The total number of entries should have reflected booked appointments and walk-ins

• Impact Question
Booking Data

- 28 unique users
  - 11 for Wrangler #1
    - This reflects Wrangler #1’s Summer fellowship
  - 17 for Wrangler #2
    - One was wrangler #2’s email, so true amount is 27

- 58 total bookings
  - 21 for Wrangler #1
  - 37 for Wrangler #2
Future Considerations

• Training
  – Lack of supervisor subject knowledge
    • Shadowing
    • Watching online tutorials

• Recruiting
  • Must hire technical experts
  • Train them to have consultations and teach
  • Opposite of most other student library positions
  – Employee Turnover
    • Length of tenure heavily dependent on year of program
    • Near-continuous training effort
Future Considerations (cont.)

• Budgetary Concerns
  – Cuts to student employee budgets
    • Limited hours
  – Competition with other local employers for technically skilled students
Future Considerations (cont.)

• Assessment
  – Better Record Keeping
    • Wrangler Questionnaire form
      – Simplify
      – Train to always fill out
    • Booking stats
      – Don’t tell whole picture
        » No shows, walk-ins not captured
  – Survey Reliability
    • Response Rate
      – Fake email to get more appointments
Thank You
Questions?

Mitch Winterman, MLIS
775-682-5510
mwinterman@unr.edu