Revisiting Low-Tech in a High Throughput Environment

Glen Ritchie[1,2], Nothabo Dube[1], and Irish Pabuayon[1]
[1]Texas Tech University Plant and Soil Science
[2]Texas A&M AgriLife Research

Abstract:
Data collection and management have changed substantially over the past few decades. High throughput data collection and analysis are now possible through the use of faster processing, smaller data collection platforms, and improved software for analysis. However, it is important to remember that agricultural research is still a human endeavor. This poster describes some of the principles and skills that are still important to remember as we move into high-throughput environments. These principles include research planning, remembering the role of data, auditing, analysis, and interpretation of results.

Introduction:
In May, 1997, a supercomputer called Deep Blue defeated world chess champion Gary Kasparov 2:1 in 6 matches. At the time, Deep Blue was one of the most powerful computers in the world. Twenty years later, Deep Bue's processing power is dwarfed by that of most modern smartphones (T., N., 2014). Unfortunately, the increased processing power and reduced size of our current technologies sometimes leads students and experienced researchers alike to skip steps that are important to robust research. The following principles are important to remember in the current data-centric environment (Fig. 1):

- Research planning
- Remembering the role of data
- Data auditing
- Analysis is not a one-step process
- Interpretation is still the end-game

Fig. 1. Agronomic research requires an integration of planning, collection, and interpretation.

Research Planning
Well-defined objectives have the following value in research:

- Minimizing holes in data collection and hypothesis testing
- Minimizing data collection that is not pertinent to objectives
- Identification of project limitations and estimation of treatment, replication, and room requirements (Fig. 2)

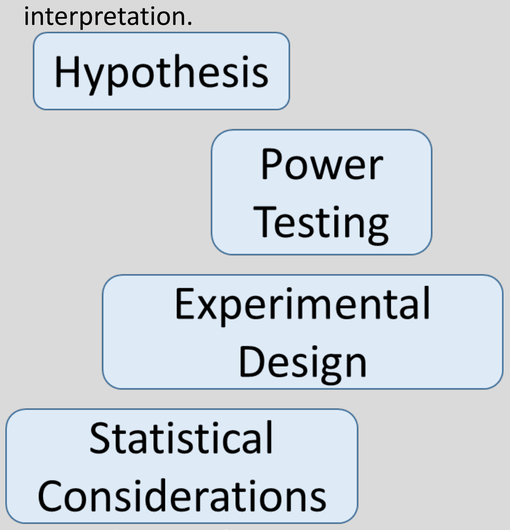


Fig. 2. Planning of experiment.

Remembering the Role of Data
Graduate students often consider data the end goal of field research. Many measurements are rapid and can generate thousands, millions, or more data points over the course of the season. Although datasets are critical to research, they are not the end in and of themselves.

- What is the mechanism behind the results?
- What do they mean to the hypothesis?

Data are not the research project. They are observations that lead to analysis and interpretation (Fig. 3).

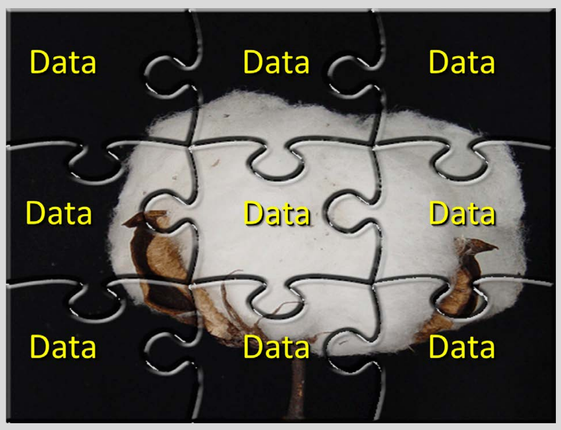


Fig. 3. If data were the whole purpose, everything would be data.

Data Auditing
Data auditing is not a new concept, but it becomes more difficult, and more important, with high quantities of information. It sometimes helps to view data by location within the field. For example, the treatments on the right seem at first glance to be good: consistency within treatments, differences among treatments, apparent blocking and randomization (Table 1). However, plotting the yields within the field design indicate border effects on both sides of the field (Table 2).
Other ways to audit data:

- Filtering for excessively large or small values
- Using GPS and other positioning verification
- In-season observations of plots that are damaged or may have been influenced by factors other than treatment.

Tables 1 and 2. Sometimes, design issues are only easy to identify within the layout.

Plot	Replicate	Treatment	Yield (kg/ha)
101	1	1	1105
102	1	2	927
103	1	3	959
104	1	4	1076
201	2	3	1058
202	2	2	899
203	2	4	878
204	2	1	1056
301	3	1	1091
302	3	4	856
303	3	2	879
304	3	3	1146
401	4	4	1073
402	4	2	986
403	4	3	925
404	4	1	1107

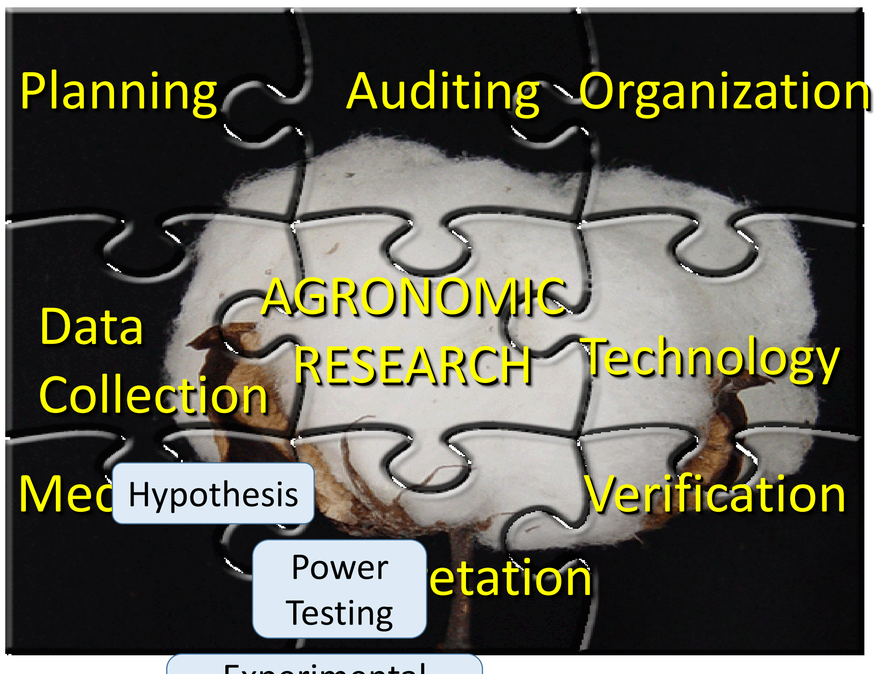
Trt 4-1073 kg/ha	Trt 2-986 kg/ha	Trt 3-925 kg/ha	Trt 1-1107 kg/ha
Trt 1-1091 kg/ha	Trt 4-856 kg/ha	Trt 2-879 kg/ha	Trt 3-1146 kg/ha
Trt 3-1058 kg/ha	Trt 2-899 kg/ha	Trt 4-878 kg/ha	Trt 1-1056 kg/ha
Trt 1-1105 kg/ha	Trt 2-927 kg/ha	Trt 3-959 kg/ha	Trt 4-1076 kg/ha

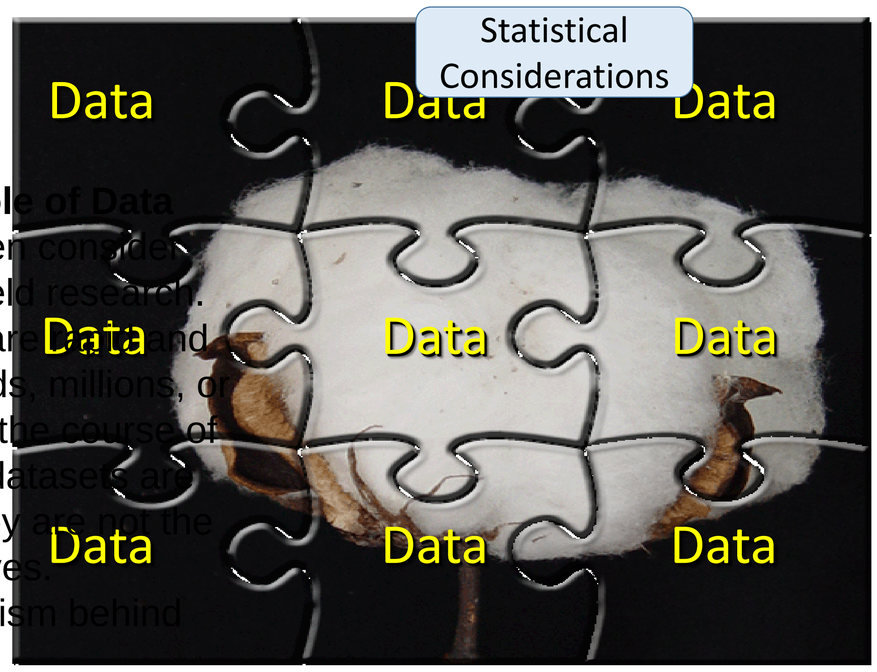
Data Analysis
Data analysis is not a one-step process. Mean separation and regression require identification of normality. ANOVA and other mean separation methods also require a fundamental understanding of model effects and model structure. For example, proc GLM, a popular model for ANOVA analysis in SAS, does not support random effects in its model structure and requires post hoc analysis. However, several researchers still use proc GLM in situations where random effects are present.
It is also important to remember that the reason why results are significant is as important as the significance itself.

Interpretation
As a technical editor and as a graduate student advisor, the biggest and most identifiable issues I see with manuscripts are based on the results, discussion, and conclusions. A weak discussion or conclusion section suggests that the authors have not done their homework in designing the research, or that they have a fundamental weakness in understanding their findings and the importance of their work.
A strong discussion compares and extends on previous researchers' work.
Strong conclusions show a strong understanding of the problems addressed through the research and the vision of additional work that can be done to extend the research.

Reference
T., N. 2014. A modern smartphone or a vintage supercomputer: which is more powerful? Retrieved October 21, 2017, from https://www.phonearena.com/news/A-modern-smartphone-or-a-vintage-supercomputer-which-is-more-powerful_id57149

Plot 401-1073 kg/ha	Plot 402-986 kg/ha	Plot 403-925 kg/ha	Plot 404-1107 kg/ha
Plot 301-1091 kg/ha	Plot 302-856 kg/ha	Plot 303-879 kg/ha	Plot 304-1146 kg/ha
Plot 201-1058 kg/ha	Plot 202-899 kg/ha	Plot 203-878 kg/ha	Plot 204-1056 kg/ha
Plot 101-1105 kg/ha	Plot 102-927 kg/ha	Plot 103-959 kg/ha	Plot 104-1076 kg/ha





Remembering the Role of Data

Graduate students often consider data the end goal of field research. Many measurements are taken and can generate thousands, millions, or more data points over the course of the season. Although datasets are critical to research, they are not the end in and of themselves.

- What is the mechanism behind the results?
- What do they mean to the hypothesis?