



Data Science and collaboration skills for synthesis science

A training workshop for SNAPP and NCEAS postdoctoral associates, held at NCEAS, UCSB, Jan 18-21, 2020

Sponsored by:




SNAPP

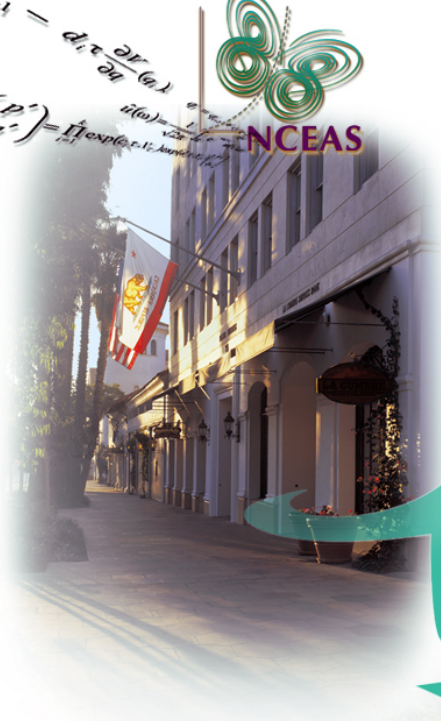
Science for Nature and People
PARTNERSHIP

Welcome to NCEAS!

We are turning 25 this year!!

$$P_i \equiv P_{i-1} = d_i \tau \frac{\partial V}{\partial q}(q_i)$$
$$(q') = \{q'_{\text{exploit}}, q'_{\text{conserv}}, q'_{\text{other}}\}$$


The logo for the National Center for Ecological and Evolutionary Synthesis (NCEAS) is positioned above the word "NCEAS" in a purple, serif font. The logo itself is a stylized green butterfly or flower-like shape with multiple loops.





NCEAS

National Center for Ecological Analysis and Synthesis

Synthesis Science Research

- Environmental, ecological, conservation, social, and earth sciences

Informatics

- Research and development of new tools/solutions
- Data Science support

Learning Hub - Training

- Computational and collaboration skills

Logistics

- Wireless password is... on the boards
- Bathroom code is... on the boards
- Snacks, water, coffee & tea in lounge (upstairs)
- Remote participants:
Reimbursements == keep receipts!!!
- Lunch will be provided in the lounge the 3 first days

A Challenge for Synthesis

It is often said that 80% of data analysis is spent on the process of cleaning and preparing the data *

* Wickham 2014. *Tidy Data*. J. Stat Soft 59(10):

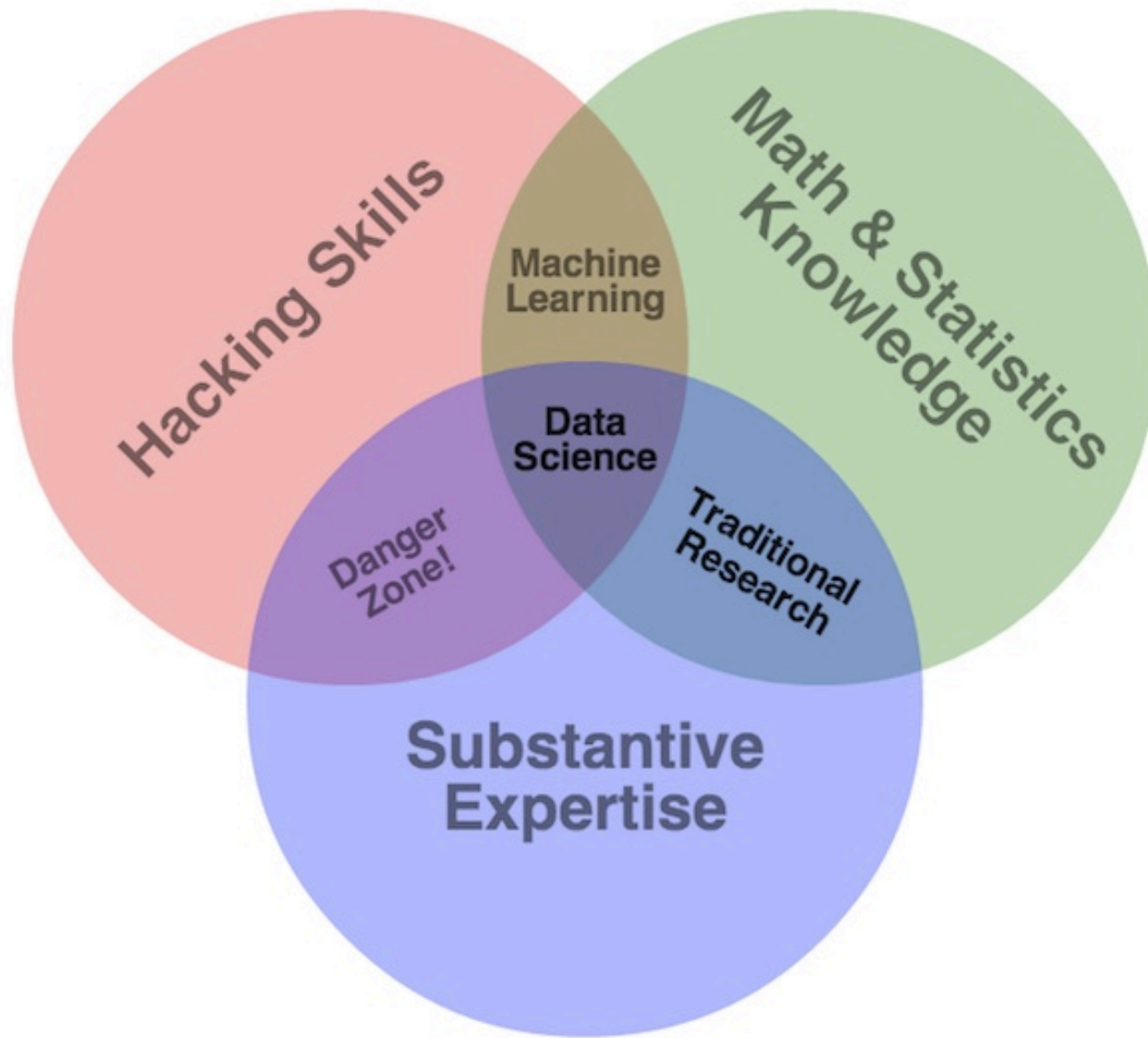


Image Attribution: Drew Conway

<http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>

The New Age of “(Data) Science”

“Open Science takes over:

In principle, the purpose of scientific publication is to enable ***reproducibility*** of research findings. For centuries, computational results and data analyses have been referred to in scientific publication, but typically only have given readers a hint of the full complexity of the data analysis being described...

To meet the original goal of scientific publication, one should share the underlying code and data.” *

* Donoho, D. 2015. “50 years of Data Science”, presented at Tukey Centennial Workshop.

Workshop Goals

- Instill understanding of *collaborative science best practices* within SNAPP and NCEAS postdocs communities
- Empower participants to adopt *computing best practices* in their scientific workflows and data management
- Foster potential collaboration and enhanced knowledge-sharing across Working Groups and Projects
- Familiarize participants in the use of (NCEAS) analytical server
- Reinforce commitment to Reproducible Science

Workshop Expectations

- Participate!
- Be kind!
- Learn some new technology concepts and tools!
- Create connections
- Don't be afraid to ask questions!
- If you already know a topic, help out your colleagues who might not
- Have fun!!