

Introduction to

DATA SCIENCE WORKSHOP SERIES

Learning the Tools of the Trade

With an increase in demand for the use of data science tools and workflows in the University, the KAUST Visualization Core Laboratory (KVL) has organized a series of “Introduction to Data Science Workshops” to equip the research community with core data science tools, and to enable future data science applications at KAUST.

Through this series of workshops, KVL aims to complement its existing training series on data visualization, and also advance the state-of-the-art in both data science and visualization by providing advanced facilities, training, services, and consulting to the KAUST community and the Kingdom.

Identifying Core Competencies for Data Science

According to a recent O'Reilly Data Science Survey, most data scientists use multiple programming languages on a daily basis to solve their data science problems. The top three programming languages used by data scientists are SQL, Python, and Bash. The ability to share and reproduce data science workflows is critical, whether the workflows are providing decision support in industrial applications, or generating novel insights from scientific data. Core tools for facilitating reproducible data science workflows are version control tools such as Git or virtual environment tools such as Conda.

Building Data Science Capacity at KAUST

KVL has organized a series of Introduction to Data Science workshops (<https://wiki.vis.kaust.edu.sa/training>) to build capacity in the core data science tools and enable future data science applications at KAUST.

- Introduction to Shell for (Data) Scientists: September 15, 2020
- Introduction to Conda for (Data) Scientists: September 22 and November 03, 2020
- Introduction to Python for Data Science: September 29 and October 20, 2020
- Introduction to Version Control using Git for (Data) Scientists: October 6, 2020
- Introduction to SQL for Data Science: October 13, 2020

The core workshop material largely follows a curriculum developed by Software and Data Carpentry, two global nonprofit organizations that teach foundational coding and data science skills to researchers worldwide. The curriculum will be offered every Fall and Spring semester in its entirety in order to provide KAUST students, post-docs, staff, and researchers with an opportunity to develop their skills in these core data science tools.

KAUST Core Labs will offer a Certificate of Completion to those learners who complete the core Introduction to Data Science curriculum.

Helping to Advance the State-of-the-Art in Data Science at KAUST

In addition to building capacity in core data science tools, KVL and KAUST Supercomputing Core Laboratory (KSL) are planning to offer additional advanced training courses in tools used in state-of-the-art data science applications with a particular focus on enabling data science with GPUs.