

Unit 2: Welcome to your home

Why UNIX?

Lesson 11

Derek Ruths

Overview of unit

Objectives:

- Fundamentals of working in UNIX
- Fundamentals of working in the cloud

- | | |
|--------------------------------|----------------------------|
| 1. Why UNIX? | 7. Installing python |
| 2. Why AWS and the Cloud? | 8. Intermediate UNIX tools |
| 3. Setting up a UNIX VM | 9. Working remote |
| 4. SSH | 10. Text editors |
| 5. Key-based Auth | 11. Advanced customization |
| 6. Basic UNIX commands & admin | 12. HW 2 |

Lesson overview

Objectives

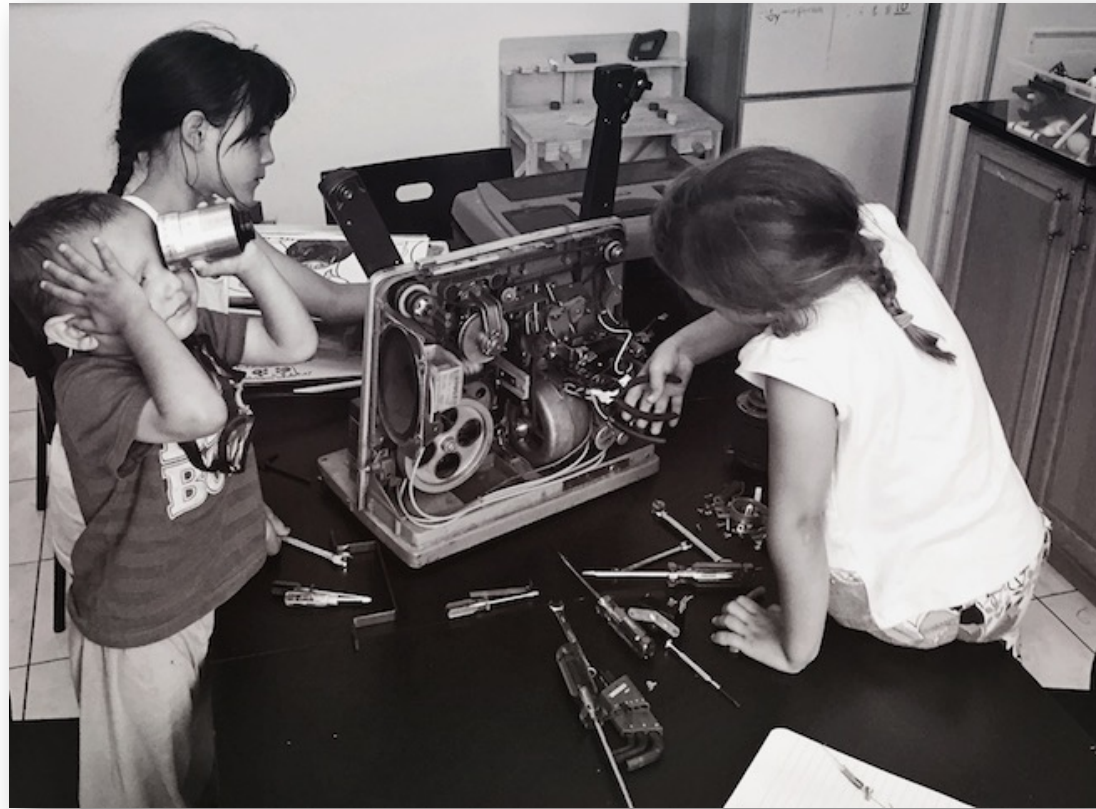
- Understand why UNIX is the standard platform for data science

Outline

- Working with data
- A super brief history on UNIX



Working with data



Introduction to Data Science
Derek Ruths

Working with data

Exploration: initially, data is an unknown object

- You will need to apply all kinds of tools in no predictable order
- Python is great, but it has serious limitations, inefficient, and clunky

Scale: data can be BIG

- You need to be able to quickly slice a piece of it to play with

Visualization: it's about story telling, not initial analysis

- With dirty data, visualization won't be your first stop

A super brief history of UNIX

- UNIX was born in the early 1970's before graphics were a “thing”.
- A strong philosophy evolved around small, modular tools that could work well together: [Peter H. Salus]
 - Write programs to do one thing and do it well
 - Write programs to work together
 - Write programs to handle text streams, because that is a universal interface.
- This describes EXACTLY what we need when we work with data.

Working with data

1. Write programs to do one thing and do it well
2. Write programs to work together
3. Write programs to handle text streams, because that is a universal interface.

Exploration: initially, data is an unknown object

- You will need to apply all kinds of tools in no predictable order
- Python is great, but it has serious limitations, inefficient, and clunky

Scale: data can be BIG

- You need to be able to quickly slice a piece of it to play with

Visualization: it's about story telling, not initial analysis

- With dirty data, visualization won't often be your first stop

Lesson wrap-up

Takeaways

- UNIX philosophy – modularity, integration, text
- Data science agility maps tightly to the UNIX philosophy

Up next

- Computing in the cloud

WSL