Introduction to Python for Business Analytics and Visualization

Presenter: Steve Baskauf
steve.baskauf@vanderbilt.edu

DISC DIGITAL SCHOLARSHIP AND COMMUNICATIONS

Jean & Alexander Heard LIBRARIES

Remote Support for Teaching and Research Needs

- Access to digital collections 24/7
- Skype consultations with your subject librarian
- Ask a Librarian: an easy way to submit a question via email
- Live chat available from the Library home page

Jean & Alexander Heard
LIBRARIES

NEED HELP? ASK A LIBRARIAN!

https://www.library.vanderbilt.edu/ask-librarian.php

Jean & Alexander Heard
LIBRARIES

Digital Scholarship and Communications Office (DiSC)

- Unit of the Vanderbilt Libraries
- Support for data best practices (DMP tool, repositories), GIS, copyright, Linked Data (including Wikidata), tools (GitHub, ORCID, Open Science Framework, etc.), and Open Access publishing.
- Offers on-demand educational programming, consultations, web resources
- Currently offering lessons on Python, R, and GIS
- More online at: **vanderbi.lt/disc**
- Email: disc@vanderbilt.edu

Introduction

DISC DIGITAL SCHOLARSHIP AND COMMUNICATIONS

Jean & Alexander Heard LIBRARIES

Topics for today

1. Introduction including philosophy
2. Lesson structure
3. Software installation
4. Questions

About me

- High School teacher (10 years)
- Ph.D. biologist (20 years)
- Data science and data curation specialist (2 years)

- I believe in making the cake before starting on the icing.
- I'm more interested in "production" than "marketing" (see these slides and the website).

What is "CodeGraf"

- A new paradigm born:
  - after 1.5 years of teaching Python and R lessons
  - after half a semester of remote teaching with Zoom
- What I learned:
  - one-hour class sessions (and recordings) are too long
  - not enough time for practice between introduction of topics
  - what learners want and need are all different
- With CodeGraf:
  - work at your own pace
  - lessons in small chunks that can be skipped or repeated
  - "choose your own adventure" style (eventually)
- Landing page: **`vanderbi.lt/codegraf`**

About Python

- Python 2 vs. 3
  - Python 2 reached "end of life" on Jan. 1 this year.
  - Many Python 2 code examples lurk on the web.
- Python is a language, not an application
  - must learn to write scripts, not operate software
  - there is no single way to run it
  - paintball vs. Arabic poetry (cf. viz with Tableau)
- Python can be used many ways
  - linear scripting
  - procedural coding
  - web applications, running robots, AWS lambdas, etc.
- What does "I know Python" mean?

Lesson structure

DISC DIGITAL SCHOLARSHIP AND COMMUNICATIONS

Jean & Alexander Heard LIBRARIES

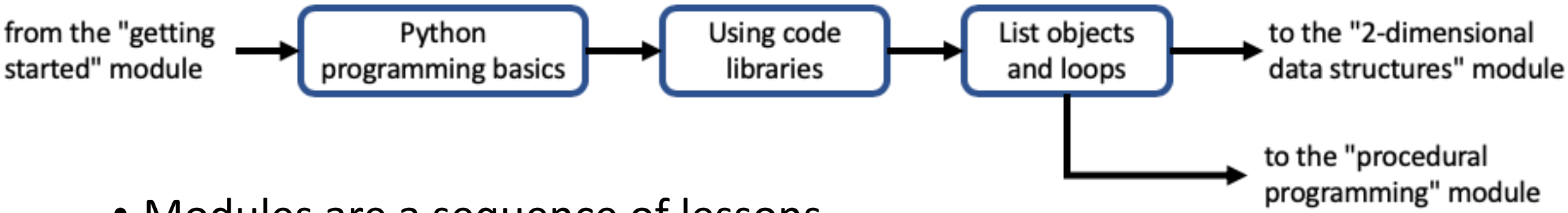
The big picture (CodeGraf "path")

1. Choose and install a Python environment (not to be underestimated)
2. Learn basic Python terminology, syntax, and structure (first week)
   - Equivalent of learning script, word construction, and grammar for Arabic poetry
3. Learn how to write scripts for:
   - NumPy array structure and file I/O; "tidy data"
   - Wrangling two-dimensional data ("spreadsheets")
   - Creating graphs ("data viz")
   - Some basic stats and analysis

Video-watching vs. practice

- In early lessons, nearly all "work" is watching videos.
  - "Why aren't we doing more practice?"
  - Since Python is a language, we can't "speak it" immediately.
  - Try modifying Jupyter notebook code cells to see what happens.
- Towards the end of the "Introductory coding" module, the opportunity for coding practice will explode.
- The last "script writing" track will provide more opportunities to modify template scripts ("cookbook")

Lessons structure



- Modules are a sequence of lessons
  - Lessons need to be done in order
  - Each lesson is on a separate web page
- Lessons are a series of topics
  - The topics are mostly sequential
  - Each topic has a video
  - Many topics have text, code examples, and links
  - Each lesson should have some kind of "practice" at the end

Can I skip lessons or topics?

- Yes, if you think you already know it – you can always go back
- Some the content in the "Programming environments" and "Installing an environment" lessons provide useful background, not just installation instructions.

Software installation

DISC DIGITAL SCHOLARSHIP AND COMMUNICATIONS

Jean & Alexander Heard LIBRARIES

What software do I need to install?

- You should install the minimum software required to do the job.
- The lessons are all based on Jupyter notebooks. Options:
  - Google Colab (cloud-based) Best option for beginners.
  - Microsoft Azure (cloud-based)
  - local installation (with or without Anaconda) Best option for tech-oriented users.
- The entire first module revolves around understanding the options and choosing.

Jupyter notebook demo

- Local installation
- Colab notebook

Questions

DISC DIGITAL SCHOLARSHIP AND COMMUNICATIONS

Jean & Alexander Heard LIBRARIES

How fast should I be working?

- No single answer, but probably finish the first two modules by the end of the first week.
- Technically advanced participants may catch up to my lesson prep.
  - I'm updating pages "live", so they will evolve over time
  - Hope to add more links to resources and practice examples.
- Next week's session is virtual "office hours"

How to contact me

- Feel free to email brief questions to me at steve.baskauf@vanderbilt.edu
- Set up a Zoom/Teams/FaceTime live meeting for more complicated questions
- Bring up interesting/broad questions at the next session (even if answered)

- I can't keep up with Slack/Twitter/Facebook/Teams messaging/etc.