

Expanding Graduate Students Data Literacy

Ximin Mi, Data Visualization Librarian

Jay Forrest, Assistant Dean for Content Strategy and Development

20 May 2022

Data Literacy

Analytical Skills

- Information literacy
- Methods - Basic
 - Regression
 - Classification
 - Visualization
- Methods – Advanced
 - Other Statistical Techniques / Econometrics
 - Machine Learning/AI
 - Simulation

Programming Skills

- Data Wrangling
 - OpenRefine
 - SQL
- Data Analysis
 - R
 - Python
- Data Visualization
 - Tableau
 - D3
- Techniques
 - Text Mining
 - Bibliometrics
 - Big Data

Domain Knowledge

- Reason
- Data Intuition

Data Literacy: Where Students Learn

Analytical Skills

- Information literacy
- Methods - Basic
 - Undergraduate Statistics Course
 - Graduate Methods Course
- Methods – Advanced
 - Self-study

Programming Skills

- Flavor of Graduate Methods Course
- Collections
 - LinkedIn Learning
 - O'Reilly Books and Media
 - Sage Research Methods
 - Data Camp
- **Library/Campus Instruction**

Domain Knowledge

- Courses in Program
- Individual Research

Data Literacy at the GT Library

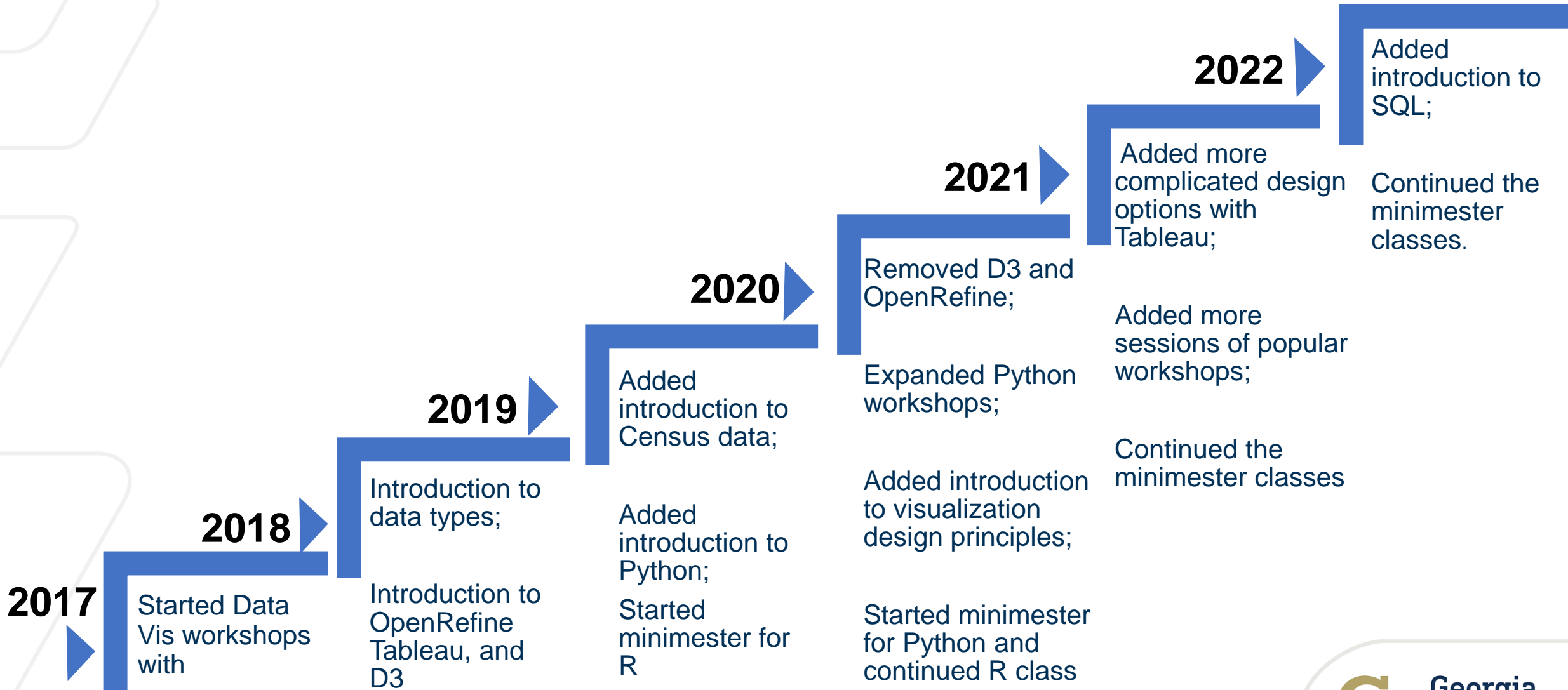
Democratization of Data Literacy

Offer everyone regardless of background (or location) an opportunity to learn tools (and basic techniques).

Data Literacy Support throughout Research Lifecycle

- Data Cleaning
 - Importing and preparing data for next steps
- Data Analysis
 - Using software tools to analyze data
- Data Visualization
 - Using software tools to visualize data

Data Literacy Instruction at GT Library



Minimester Class: Python for Visualization

Understand:

- basic data-driven research methods
- basic Python logic and functions
- data structure
- how Python interacts with different files / data types

Conduct:

- statistical analysis with Python
- visualization with Python

Communicate:

- the final product with an audience

Minimester (15 hours over 5 weeks)

Open Science with R

- Advance comprehension about Open Science principles including: data sharing, open data tools, and open reproducible research.
- Identify how to input data from a variety of formats into R (readr)
- Describe how and why to transform data in R (dplyr, tidyr, stringr)
- Explain how to visualize data with R (ggplot2)
- Be able to conduct Exploratory Data Analysis in R
- Be able to create basic statistical models in R (, lm, rpart)
- Be able to communicate and share data (RMarkdown language)

Data Science Boot Camp

- Advance comprehension about Open Science principles including: data sharing, open data tools, and open reproducible research.
- Collect data with survey and api tools.
- Understand how to input structure and unstructured data into R and Python
- Understand why and how to clean and transform data in R, Python, and OpenRefine
- Be able to conduct Exploratory Data Analysis in R and Python
- Understand how to visualize data using ggplot and plotly
- Fundamentals of GIS with QGIS

R (Social Sciences)

- 2017
 - Introduction to R
- 2018
 - Introduction to R (subject specific)
 - Intermediate R – Data Wrangling
 - Intermediate R – Text Mining
- 2019
 - Intermediate R – Visualization
 - Intermediate R – Spatial Analysis
 - R Minimester
- 2020
 - Intermediate R – Bibliometrics
 - Intermediate R – R Markdown
 - Social Science Data Boot Camp
 - R/Python for Public Policy
- Intro
 - Importing Data
 - Regression Analysis (single and multivariate)
 - Classification (Decision Trees)
- Intermediate
 - Data Wrangling / Data Transformation
 - Join, Merge, add columns/rows, summarize
 - Visualization
 - Grammar of Graphics + ggplot
 - Bibliometrics
 - Bibliometrix
 - R Markdown

Transition to Online (2020)

Freedom to Succeed/Fail

Tools

- Bluejeans Events
- EventBrite

Offer frequently and often

- Scale offerings as demand increases/decreases and as instructor time allows.

In-Person vs. Online

	Advantages	Disadvantages
In-Person	Pre-installed software One-on-One Troubleshooting	Limited Number of Seats Limited Access
Online	Wider Audience Larger Audience	Limited Troubleshooting Need everyone on the same platform

Online Best Practice

Cloud Environments

- Rstudio.cloud
- Google Collab
- Tableau Public & Server
- Github for content sharing

Reminder before class

Early login

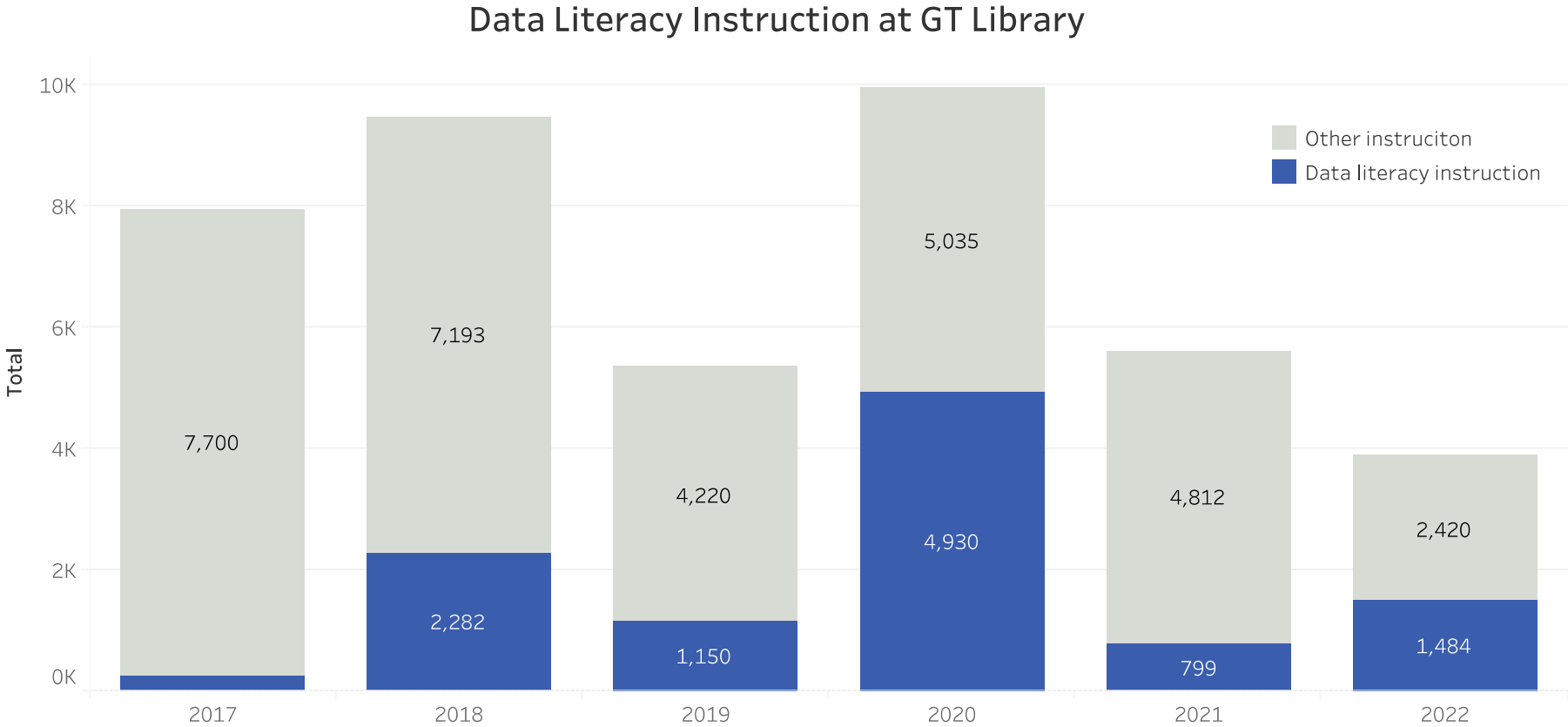
Slow demonstration of setting up the environment

Regular check-in with the online audience

Moderator in session

- Breakout with an individual student to troubleshoot or bring up to the rest of the class

Measuring the Impact: Instruction 2019-2021



Learning Together Georgia

<https://www.library.gatech.edu/learning-together-ga>

Data Analysis				
<i>Learn more about effectively evaluating data using analytical or statistical tools to discover useful information with these online courses, recordings and resources.</i>				
Corpus Linguistics (Text Mining & Text Corpora)			UGA	
Excel for Digital Humanities			UGA	
Nvivo		GSU		
OpenRefine			UGA	
Python	GT	GSU		
R	GT	GSU	UGA	
Intermediate R Studio: Data Cleaning	GT			
Intermediate R Studio: Data Visualization	GT			
Introduction to R Studio for Economics	GT			
Intermediate R Studio: Bibliometrix	GT			
SAS		GSU		
SPSS		GSU		
Stata		GSU		
Webscraping (Python & R)		GSU	UGA	
Data Visualization				
<i>Create beautiful graphical representations of data sets that easily communicate your ideas or research to a broad audience. Learn more about the tools that can assist you.</i>				
D3	GT			
Plotly	GT			
Social Explorer		GSU		

Class offerings

Beyond Instruction: Data Visualization Service

In-depth research support for tool literacy and visual design

User experience test

Use of space for project showcasing

What's next

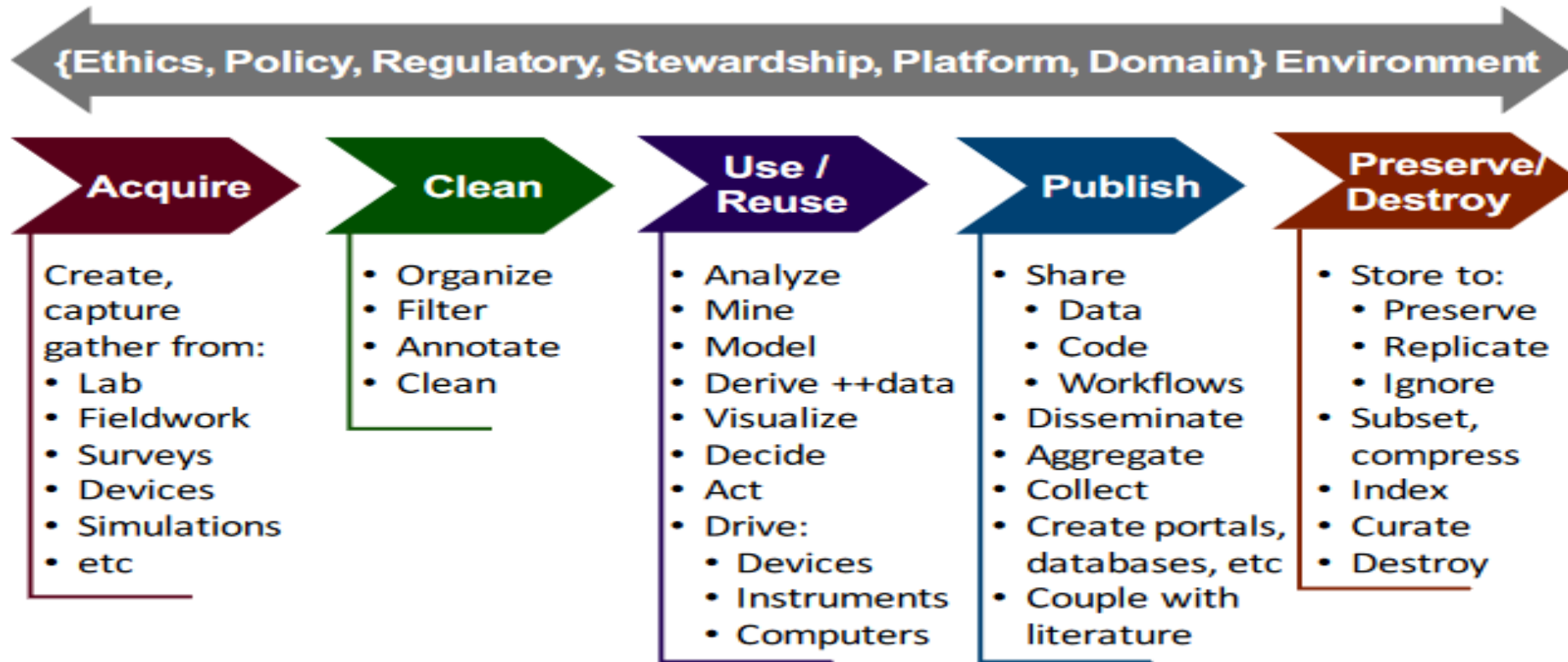


FIGURE 1: The Data Life Cycle and Surrounding Data Ecosystem