



Viya For Learners Data Repository Project Journey

May 16, 2022



Overview

Goals and Functionality



SAS Academic Programs
Transforming the Future Workforce

GLOBAL ACADEMIC PROGRAMS

**Provide Resources for
Teaching and Learning**



**Build SAS Skill Aligned with
Workforce Demand**

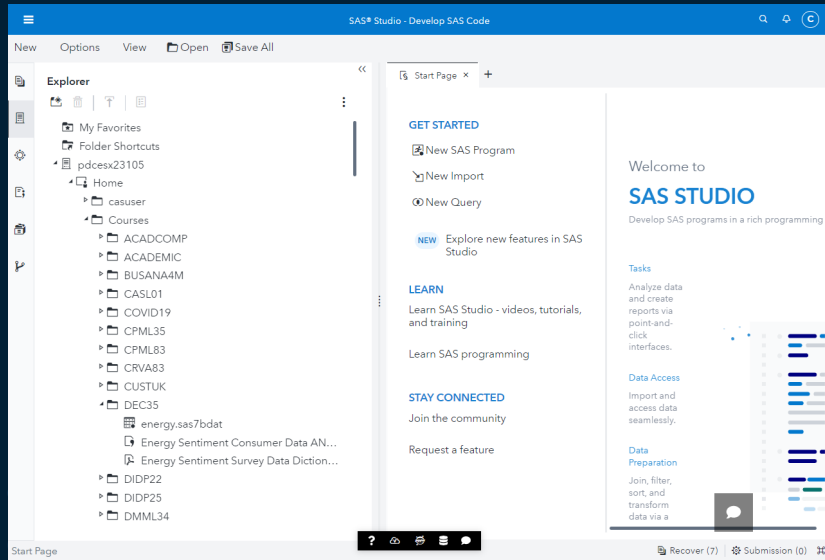


**Connect Graduates with
SAS Customers for Hiring**

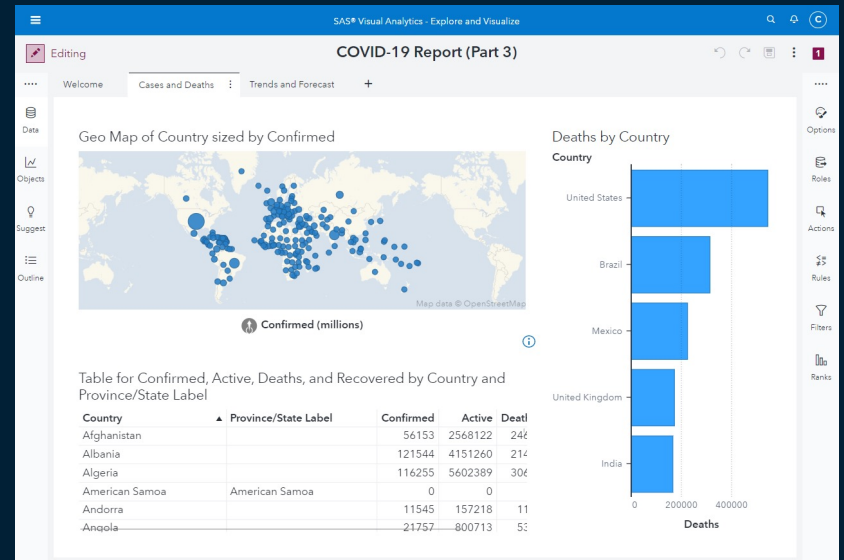


Overview

What is Viya for Learners?



The screenshot shows the SAS Studio - Develop SAS Code interface. The top menu bar includes 'New', 'Options', 'View', 'Open', and 'Save All'. The left sidebar contains an 'Explorer' pane with a file tree showing 'My Favorites', 'Folder Shortcuts', and a project named 'pdcesx23105'. The main workspace displays a 'Welcome to SAS STUDIO' message with sections for 'GET STARTED' (New SAS Program, New Import, New Query), 'LEARN' (Learn SAS Studio, Learn SAS programming), and 'STAY CONNECTED' (Join the community, Request a feature). A 'Tasks' section on the right offers to analyze data and create reports via point-and-click interfaces. The bottom status bar shows 'Recover (7)' and 'Submission (0)'.



The screenshot shows the SAS Visual Analytics - Explore and Visualize interface. The top menu bar includes 'Editing', 'Cases and Deaths', 'Trends and Forecast', and '+'. The main workspace displays a 'COVID-19 Report (Part 3)' with a 'Geo Map of Country sized by Confirmed' and a 'Deaths by Country' horizontal bar chart. Below the charts is a table for 'Confirmed, Active, Deaths, and Recovered by Country and Province/State Label'.

| Country | Province/State Label | Confirmed | Active | Death |
|----------------|----------------------|-----------|---------|-------|
| Afghanistan | | 56153 | 2568122 | 246 |
| Albania | | 121544 | 4151260 | 214 |
| Algeria | | 116255 | 5602389 | 306 |
| American Samoa | American Samoa | 0 | 0 | |
| Andorra | | 11545 | 157218 | 11 |
| Anqola | | 21757 | 800713 | 55 |

Overview

Guiding Question

What is available in Viya for Learners that can be used in course work, research, and projects?

Overview

Goals

Entries per page: 10 Search Search/sort by Dataset Name Ascending Advanced search

Apply filter

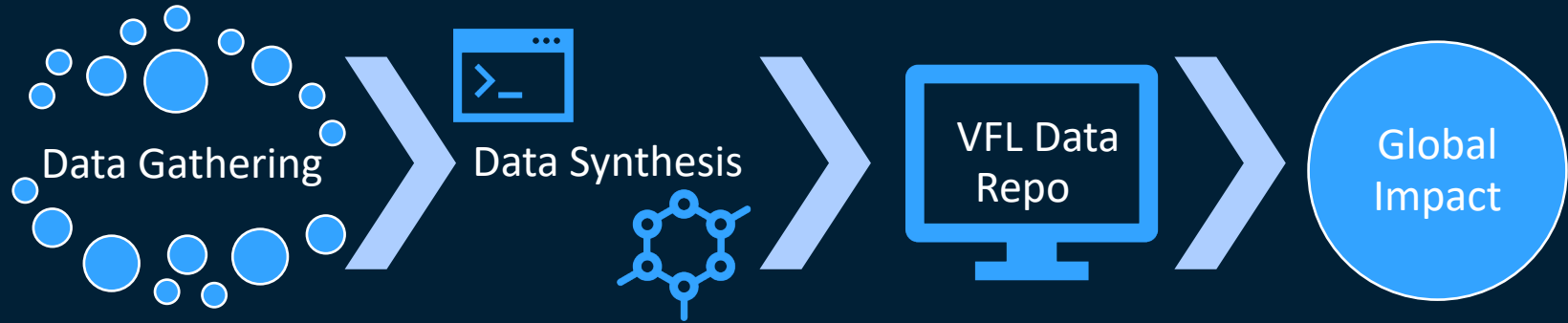
1 2 3 4 5 6 7 8 9 10 ... 21 »

| Name | Description | Industry Alignment | SAS Platform | VFL Course Folder |
|---------------------|---|--------------------|---|-------------------|
| AEEM51_TRANSACTIONS | A store is interested in determining the associations between items purchased from the Health and Beauty Aids Department and the Stationery Department. The store chose to conduct a market basket analysis of specific items purchased from these two departments, including bar soap, bows, candy bars. | Health Business | Jupyter (VFL) SAS Studio (VFL) SAS Visual Analytics (VFL) | ACADEMIC |
| ABSENTEEISM | The database was created with records of absenteeism at work from July 2007 to July 2010 at a courier company in Brazil. | Business | Jupyter (VFL) SAS Studio (VFL) SAS Visual Analytics (VFL) | ACADEMIC |
| ACACIA | This data set contains growth data on <i>Acácia negra</i> , a forest species in the Southern region of Brazil. Data include information on age, diameter, height, volume, basal area. | Environment | Jupyter (VFL) SAS Studio (VFL) SAS Visual Analytics | ACADEMIC |

- Identify Existing Shared Data
- Searchable User Interface
- Data Information Documents

Overview

Process Summary



- VFL Data Set Metadata
- Research
- MIDAS Web Scraper
- Data Upload Process Standardization

- Centralization of information
- Data Set Relationships
- Automation of Document Generation

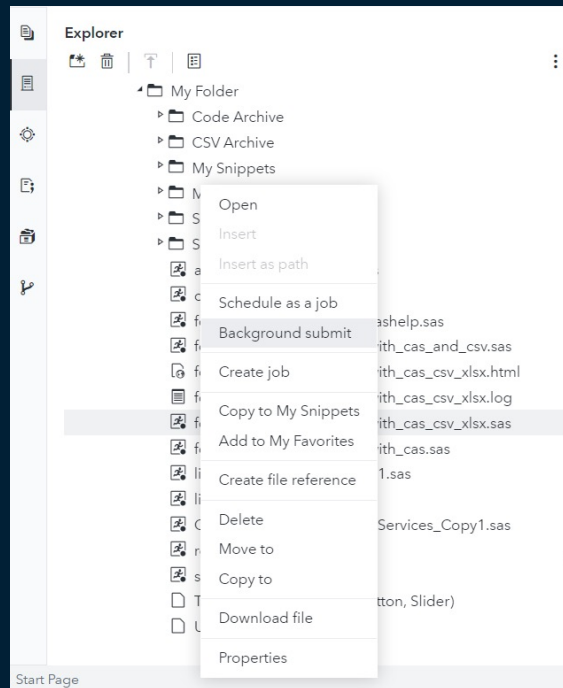
Data Gathering

Scraping, Research, Process Refinement

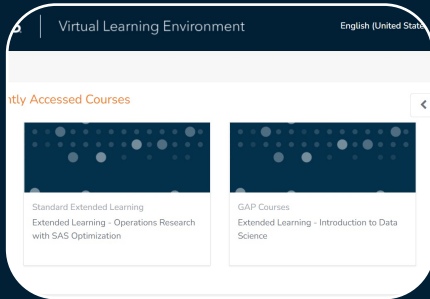
Data Gathering

VFL Data Set Metadata

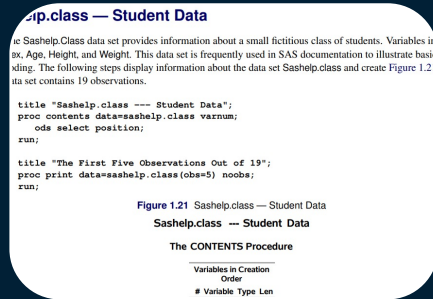
- Periodically run custom SAS program
 - Update data set list
 - Gather metadata
- Metadata
 - Name, extension, file path
 - Dimensions (variable count, observation number)
 - Date of creation/modification
 - Ability to use with cloud analytics services
 - Variable information (format, label)



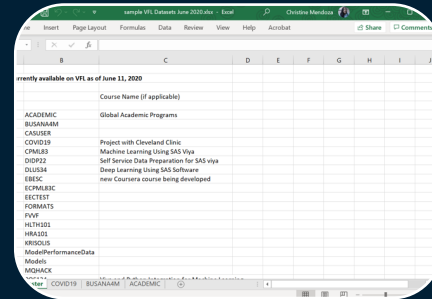
Data Gathering Research



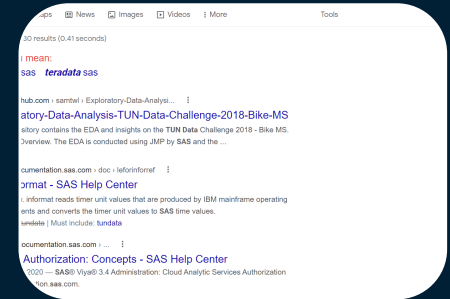
Course
Developers



SAS
Documentation



SharePoint

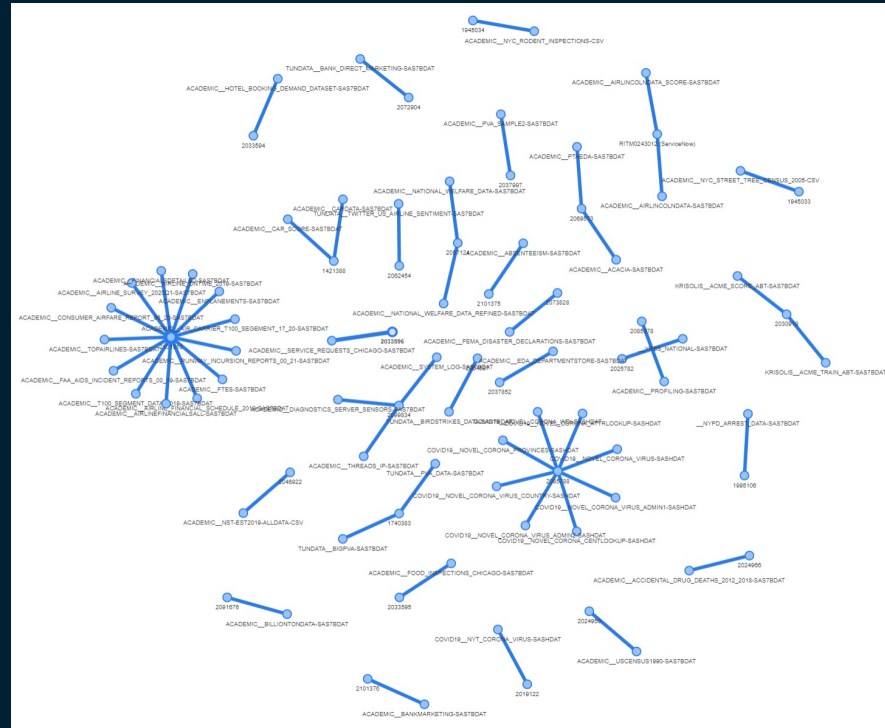


Google

Data Gathering

Upload Request Ticket Web Scraper

- Names
- Descriptions
- Source Links
- Context



Data Gathering

Data Upload Process Standardization

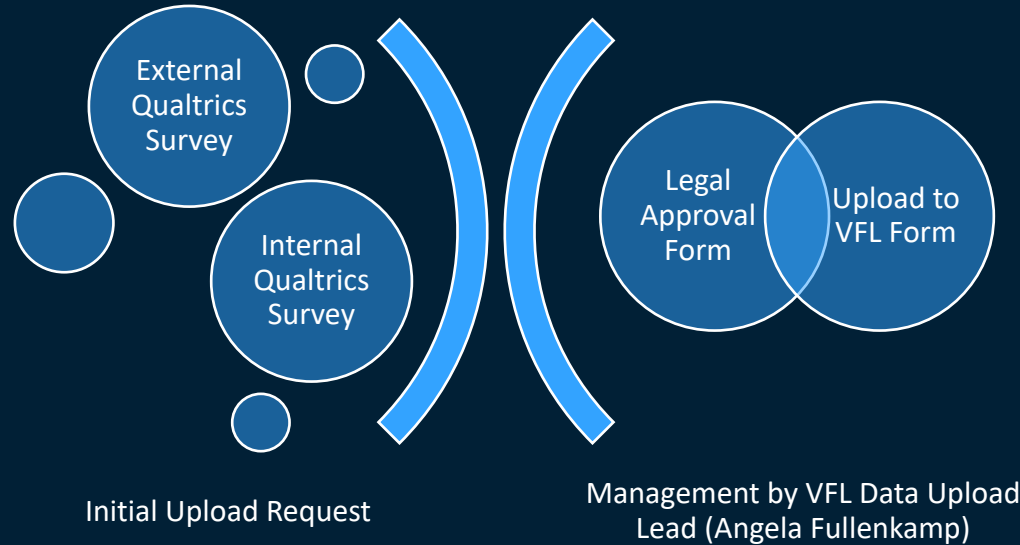
Old Process



Data Gathering

Data Upload Process Standardization

New Process



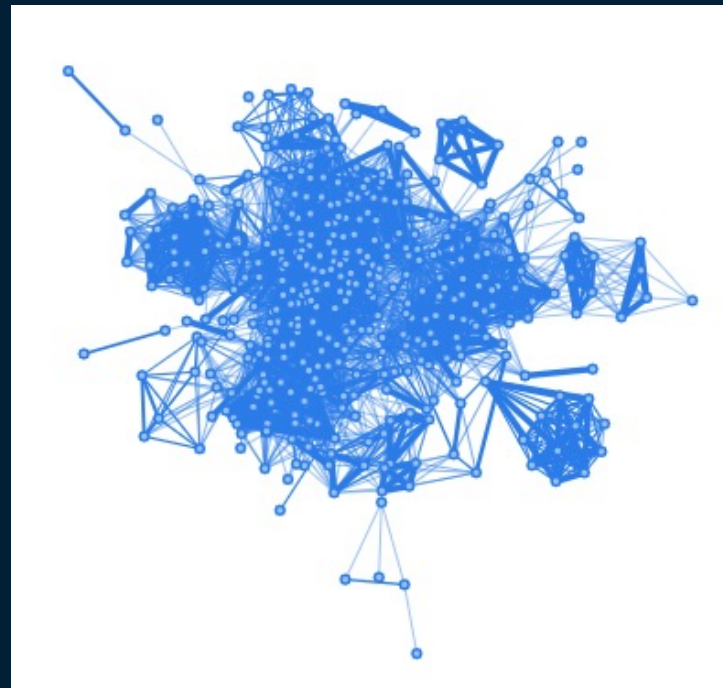
Data Synthesis

Centralization, Data Set Relationships, Document Generation

Data Synthesis

Centralization of Information

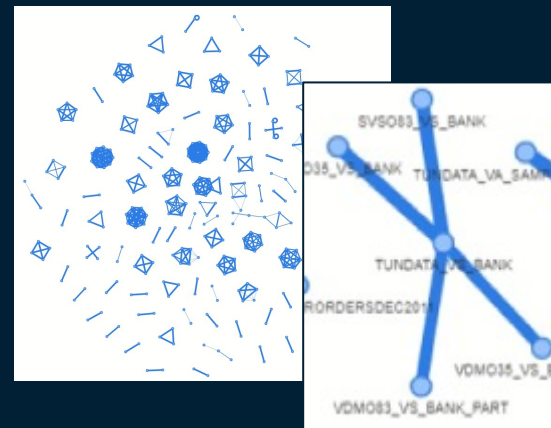
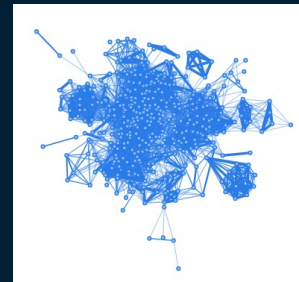
- Mock “database” on GitLab
 - Data set list from VFL data set metadata
 - Data set descriptions
 - Source links
 - Variable information
 - Industry categories
 - Information from legacy initial upload request tickets (MIDAS)



Data Synthesis

Data Set Relationships

- Problem
 - Lack of documentation of duplicates or modifications within the >500 data sets in VFL
 - Lack of links to data information on data sets
- Solution
 - Created an algorithm to relatively efficiently (<2 seconds) output a similarity metric for pairs of datasets based on their variables.
 - Displayed the results in an interactive diagram using open-source Python libraries (pandas, NetworkX, PyVis)



VFL Data Repository

Current State

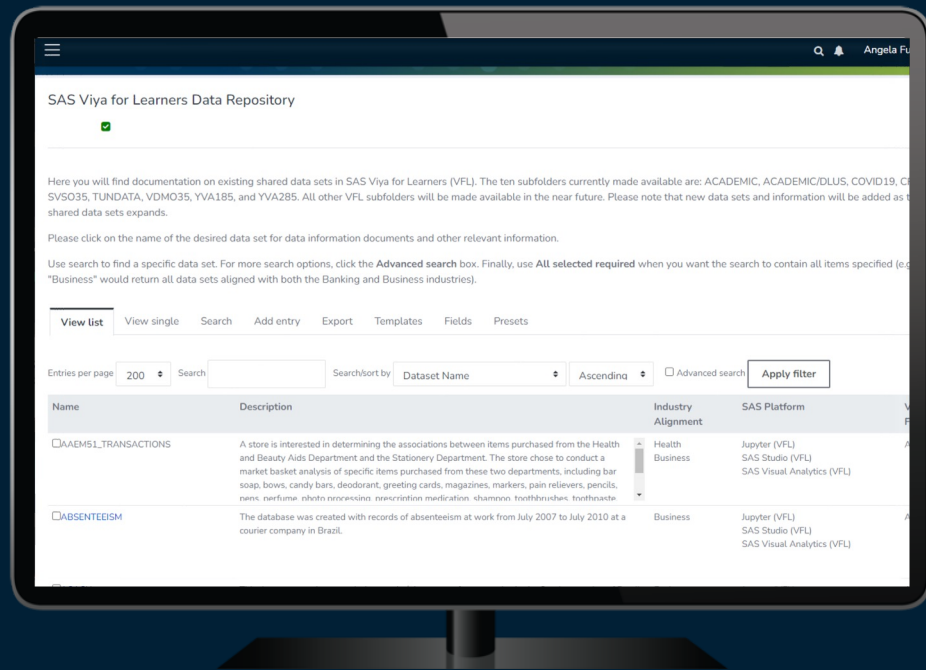
Viya for Learners Data Repository

Highlights

- Launched in January 2022
- Appx. 200 datasets catalogued

Next Milestones

- Catalog 50 datasets by end of Q2
- Catalog remaining datasets by end of Q4



Global Impact

Reach

Global Impact

Reach

Page Views

7.1K

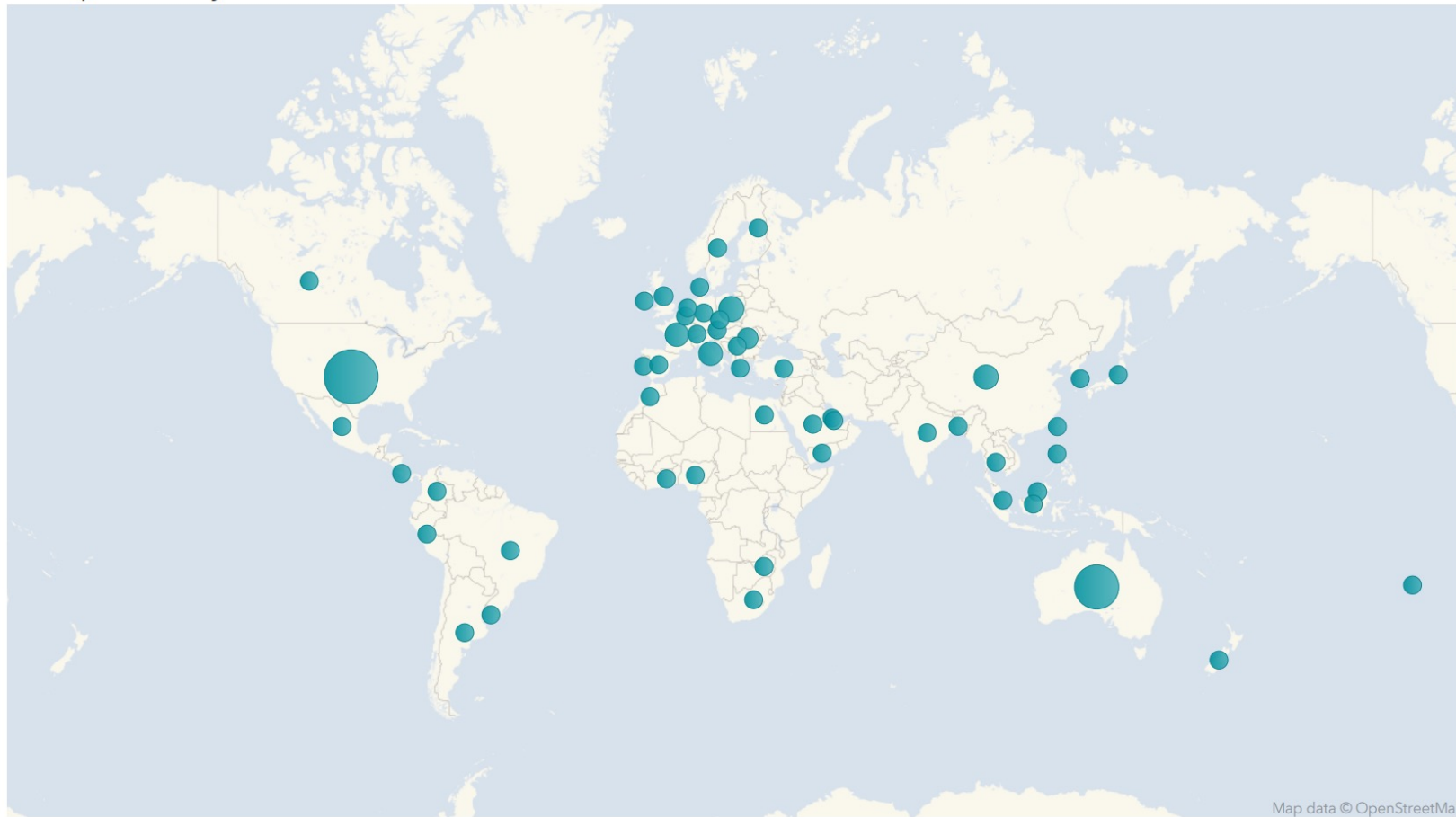
User Count

1.5K

Countries

64

Users per Country



Map data © OpenStreetMap



Frequency

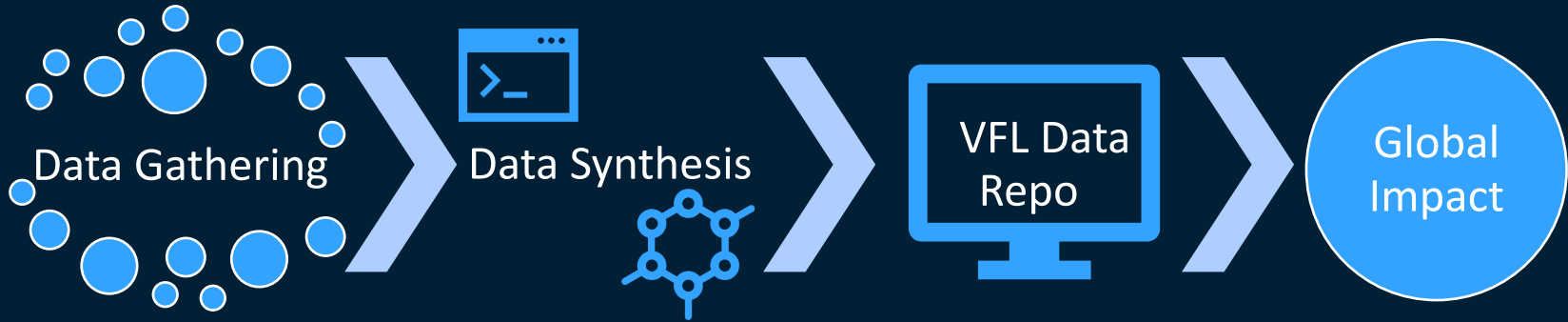


Conclusion

Summary and Acknowledgements

Conclusion

Summary



Conclusion

Acknowledgements

- Project Lead: Angela Fullenkamp
- Lincoln Groves, Lynn Letukas, James Harroun, Jim Farrell, Cheryl Cass, Aurora Pennycord-Liu, and all of my wonderful teammates, managers, and colleagues
- Course developers
- Users of VFL and the VFL data repository
- Everyone!



Thank you!

sas.com

