



Nitrogenous Fate

Susan Garcia, Aaron Lecciones, Brian Roman, Carlyn Schmidgall

CSE583 Software Development for Data Scientists
University of Washington, Seattle

Autumn 2023
13 December 2023



CSE 583 UW Fall AY 2023-24

Presentation Outline

- Problem Statement
- Proposed Solution
- Target Users
- Functional Design
- Technology Review
- Project Design
 - Github Repository
 - Project Directory Structure
 - Components
 - Scripts
 - Testing

Problem Statement

Biological oceanographers and marine biologists who study environmental samples in their microbial system experiments to characterize cellular phenotype and biochemical pathways are faced with the effects of obscuring variation when normalization is not conducted during measuring the rate of intracellular and dissolved metabolite production.

One normalization method employed is the best-matched internal standard (B-MIS) normalization which is a step-heavy process (Boysen *et al* 2018). Data visualization is also another step conducted as part of the process (Sacks *et al* 2022).

Proposed Solution

The team shall write scripts to handle data normalization and data visualization:

- Python Script for normalization of data and calculation of peak areas.
 - Panda
 - Numpy
- Jupyter Notebook for data visualization
 - Seaborn for exploratory data analysis
 - Altair to develop interactive plots

Target Users

User Characterization for NF Software

The Nitrogenous Fate Software uses metabolite data of isotopically-labeled molecules (Nitrite, Ammonia, and Urea) through incubations from samples in the Equatorial Pacific to show pathways of nitrogen within communities through simple data visualization.

=====

User	Needs and Desires	Skill Level	Use Cases
Research Scientist (Marine Biologist)	Track isotopically-labeled Nitrogen within marine communities. Desires a simple yet parameter-configurable interface with multiple options for visualization.	Highly skilled scientist who is familiar with basic data science tools	Run program as is
Research Scientist (Generic)	Use software to track defined elements in a known system. Desires a simple yet highly parameter-configurable interface with multiple options for visualization.	Highly skilled scientist who is familiar with basic data science tools and Github coding	Run program with customization
Research Scientist (administrator)	Maintain, improve, and update program as needed. Desires a simple system reporting tool for user feedback.	Highly skilled scientist who is familiar with the coding used for the program	Run and add code or debug program
Researcher	Use software but without much knowledge of underlying mechanics. Desires a way to use the software without much research or professional background in the domain	Adequately skilled researcher who might not be familiar with data science tools	Run program with assistance

Document Information: CSE583 NF Project Documentation for User Characterization

Functional Design

Version 0.1 of this software implements only the following functional design for nitrogenousfate.py:

Functional Design	Use Case	Description	Prompt
Run Program	Input Data Set	Loads data file based on run command checks for data format	May raise exceptions when encountered
	Data Set Processing	scripts runs using either customization or without customizations.	
		<i>Section I: Data Cleaning and Organization</i>	
		<i>Section II: Best Matched Internal Standard Normalization</i>	
		Data analysis encounters an error	raise Exception or display Error, exit program
		Data analysis is successful	Display: 'Analysis completed' save output files.
End Program	Terminate Program	Safely terminate program and show credits	Display thank you prompt with credits, then exit app.

Section III is handled by a Jupyter Notebook.

Document Information: CSE583 NF Project Documentation for Functional Design

Technology Review



PANDAS

pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.



NumPy

NUMPY

NumPy is a Python library that provides powerful and versatile array computations, mathematical functions, and other tools for various scientific domains



SEABORN

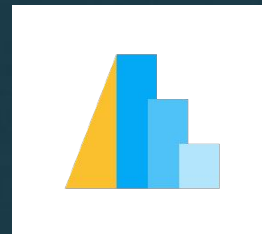
Positives:

- Standard plots out of the box
- Perfect for statistical analysis
- Fast to use for standard plots

Negatives:

- Built on top of matplotlib
- Less ability to customize

matplotlib



ALTAIR

Built on top of Vega and Vega-Lite grammars

Positives:

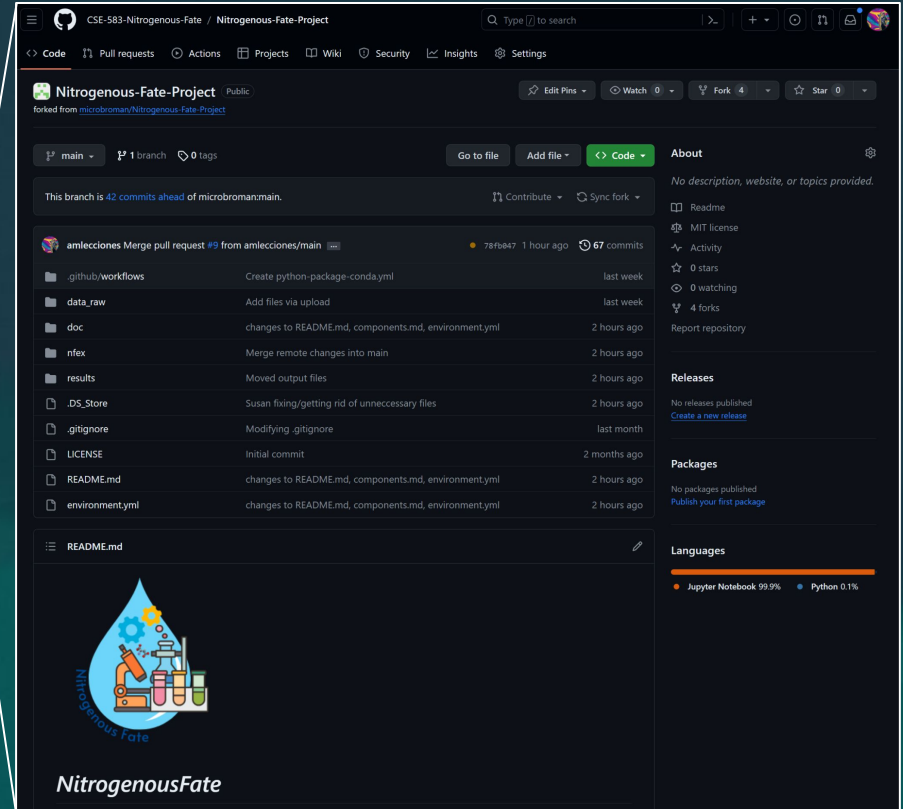
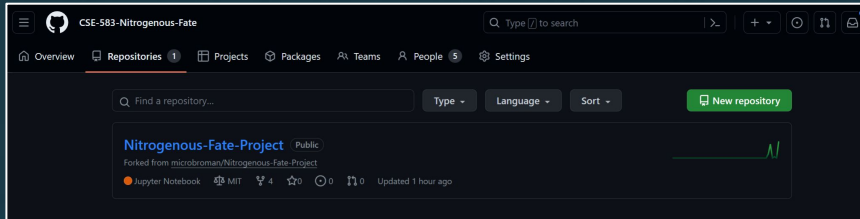
- Intuitive and structured approach to plotting
- Altair is interactive (zoom in, pan and grab, tooltips, etc)
- Flexible

Negatives:

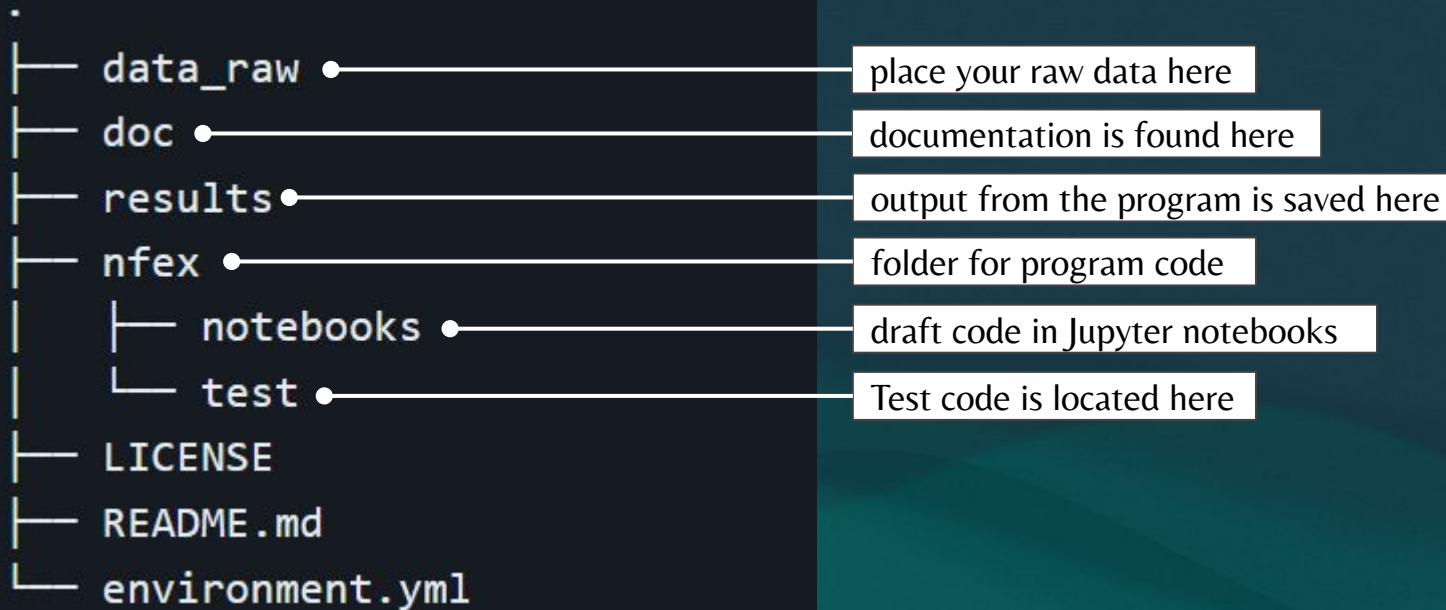
- No 3D plotting
- Not as customizable

Project Design - Github Repository

<https://github.com/orgs/CSE-583-Nitrogenous-Fate/repositories>

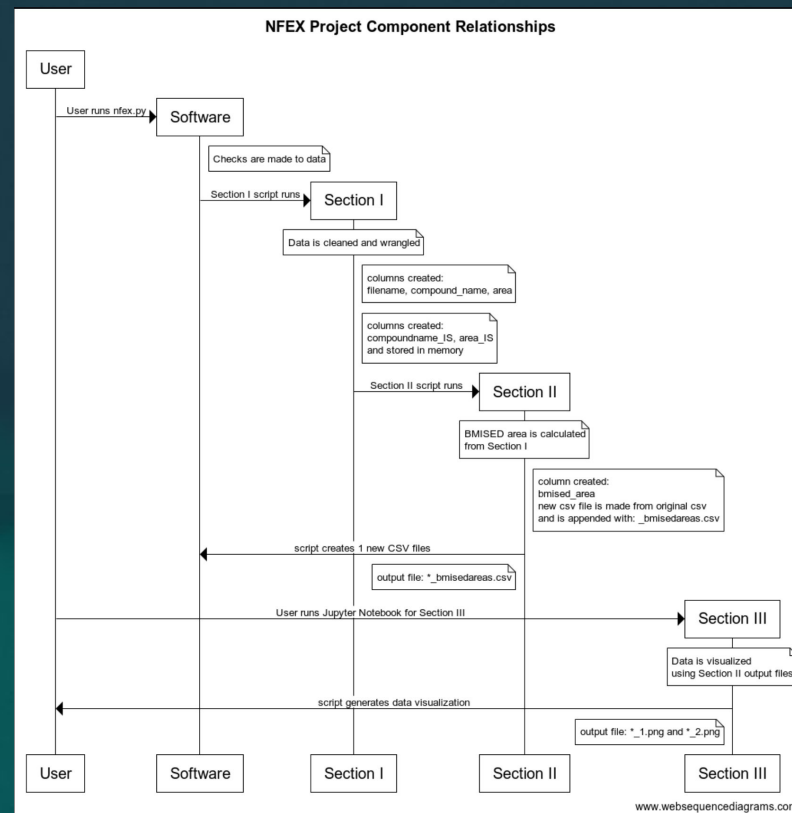


Project Design - Project Directory Structure



Project Design - Components

Name	Function	Inputs	Output	Interactions
Data Cleaning and Wrangling (Section I)	A <code>dataset</code> is loaded (either dissolved, exo or particulate, endo) and this component looks for relevant data (by column) from the data set. The script then calculates values for the following new columns which are appended to the <code>dataset</code> under a new file named <code>*_clean_areas.csv : filename, compound_name, area</code> . It then computes an internal standard which is used to normalize data for use in section II, it creates two new columns for these values: <code>compoundname_IS</code> and <code>area_IS</code> , which are appended to the <code>dataset</code> under a new file names <code>*_IS_list.csv</code> .	run <code>nfex</code> (nitrogenousfate.py) <code>Section I script</code> with data in <code>data_raw</code> folder.	component will output values for: <code>filename, compound_name, area, compoundname_IS</code> and <code>area_IS</code> .	User monitors for any errors.
BMISED Computation (Section II)	The new data in memory which now contains values in the new columns: <code>filename, compound_name, area, compoundname_IS</code> and <code>area_IS</code> is used to calculate BMISED values. These values are placed in a new column: <code>bmised_area</code> and is saved in a dataset named <code>*_bmisedareas.csv</code> .	<code>nfex Section II script</code> runs using new data stored in memory.	component will output the new dataset: <code>*_bmisedareas.csv</code> for values in the new column: <code>bmised_area</code> in <code>results</code> folder.	User monitors for any errors expects a CSV file in <code>results</code> folder.
Data Visualization (Section III)	The <code>new dataset</code> which now contains values in the new column <code>bmised_area</code> is used to generate data visualization. The Jupyter Notebook generates multiple data visualizations to choose from. The user may save them into files.	run Jupyter Notebook (XXXX.ipynb) <code>Section III script</code> with correct data in <code>results</code> folder.	component will output image files from user input in Jupyter Notebook script.	User monitors for any errors, uses Jupyter Notebook to generate and save visualization files.



Project Design - Scripts

Version 0.1

R Script

```
## Loading libraries
1 library(tidyverse)
2 library(dplyr)
3 library(janitor)
4 library(ggplot2)
5 library(cowplot)
6 library(readr)

7 # Section I: Data
8 # Cleaning and wrangling
9 export_file_name = read_csv("data/raw/NFEXgalls_LabStandards.csv")
10 # Create a transition list for
11 select(Filename, Replicate Name, compound_name, Precursor Ion Name, are
12 mutate(cpd_type = ifelse(str_detect(compound_name, "A"), "IS", "Non-IS"))
13 mutate(are = as.numeric(are))
14 mutate(are = ifelse(are == 1, "IS", "Non-IS"))
15 mutate(day = str_extract(Filename, "[0-9]{4}") %>% as.numeric())
16 mutate(day = as.numeric(day) - 1773) %>% as.numeric()
17 mutate(Filename = fct_inorder(Filename))
18 arrange(compound_name, Filename)

19 # Create a clean area list
20 filter(cpd_type == "IS")

21 standards_list_all_pos <- read_csv("data/raw/NFEXgalls_LabStandards.csv")
22 # Create a clean area list
23 filter(cpd_type == "IS")
24 filter(are == 1)
25 filter(str_detect(Filename, "A"))
26 select(compound_name, m/z = "m/z", conc_uM = "concentration_uM", cpd
27 # Create a compound type list (IS vs Non-IS)
28 right_join(clean_areas, by = "compound_name")

29 standards_list <- standards_list_all_pos %>%
30 filter(cpd_type == "Internal Standard")
```

Jupyter Notebook

```
## Section I: Data
## Cleaning and wrangling
df = pd.read_csv("data/raw/NFEXgalls_LabStandards.csv")

df.head()
```

File	Replicate Name	compound_name	Precursor Ion Name	are	Background	Area	Height	Width	Peak
1	200810_01_AutMiddBb17H2D_1	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
2	200810_01_AutMiddBb17H2D_2	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
3	200810_01_AutMiddBb17H2D_3	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
4	200810_01_AutMiddBb17H2D_4	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
5	200810_01_AutMiddBb17H2D_5	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
6	200810_01_AutMiddBb17H2D_6	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
7	200810_01_AutMiddBb17H2D_7	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
8	200810_01_AutMiddBb17H2D_8	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
9	200810_01_AutMiddBb17H2D_9	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
10	200810_01_AutMiddBb17H2D_10	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
11	200810_01_AutMiddBb17H2D_11	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
12	200810_01_AutMiddBb17H2D_12	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
13	200810_01_AutMiddBb17H2D_13	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
14	200810_01_AutMiddBb17H2D_14	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
15	200810_01_AutMiddBb17H2D_15	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
16	200810_01_AutMiddBb17H2D_16	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
17	200810_01_AutMiddBb17H2D_17	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
18	200810_01_AutMiddBb17H2D_18	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
19	200810_01_AutMiddBb17H2D_19	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
20	200810_01_AutMiddBb17H2D_20	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
21	200810_01_AutMiddBb17H2D_21	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
22	200810_01_AutMiddBb17H2D_22	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
23	200810_01_AutMiddBb17H2D_23	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
24	200810_01_AutMiddBb17H2D_24	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
25	200810_01_AutMiddBb17H2D_25	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
26	200810_01_AutMiddBb17H2D_26	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
27	200810_01_AutMiddBb17H2D_27	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
28	200810_01_AutMiddBb17H2D_28	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
29	200810_01_AutMiddBb17H2D_29	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
30	200810_01_AutMiddBb17H2D_30	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
31	200810_01_AutMiddBb17H2D_31	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
32	200810_01_AutMiddBb17H2D_32	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
33	200810_01_AutMiddBb17H2D_33	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
34	200810_01_AutMiddBb17H2D_34	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
35	200810_01_AutMiddBb17H2D_35	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
36	200810_01_AutMiddBb17H2D_36	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
37	200810_01_AutMiddBb17H2D_37	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
38	200810_01_AutMiddBb17H2D_38	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
39	200810_01_AutMiddBb17H2D_39	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
40	200810_01_AutMiddBb17H2D_40	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
41	200810_01_AutMiddBb17H2D_41	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
42	200810_01_AutMiddBb17H2D_42	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
43	200810_01_AutMiddBb17H2D_43	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
44	200810_01_AutMiddBb17H2D_44	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
45	200810_01_AutMiddBb17H2D_45	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
46	200810_01_AutMiddBb17H2D_46	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
47	200810_01_AutMiddBb17H2D_47	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
48	200810_01_AutMiddBb17H2D_48	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
49	200810_01_AutMiddBb17H2D_49	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
50	200810_01_AutMiddBb17H2D_50	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
51	200810_01_AutMiddBb17H2D_51	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
52	200810_01_AutMiddBb17H2D_52	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
53	200810_01_AutMiddBb17H2D_53	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
54	200810_01_AutMiddBb17H2D_54	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
55	200810_01_AutMiddBb17H2D_55	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
56	200810_01_AutMiddBb17H2D_56	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
57	200810_01_AutMiddBb17H2D_57	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
58	200810_01_AutMiddBb17H2D_58	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
59	200810_01_AutMiddBb17H2D_59	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
60	200810_01_AutMiddBb17H2D_60	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
61	200810_01_AutMiddBb17H2D_61	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
62	200810_01_AutMiddBb17H2D_62	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
63	200810_01_AutMiddBb17H2D_63	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
64	200810_01_AutMiddBb17H2D_64	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
65	200810_01_AutMiddBb17H2D_65	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
66	200810_01_AutMiddBb17H2D_66	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
67	200810_01_AutMiddBb17H2D_67	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
68	200810_01_AutMiddBb17H2D_68	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
69	200810_01_AutMiddBb17H2D_69	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
70	200810_01_AutMiddBb17H2D_70	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
71	200810_01_AutMiddBb17H2D_71	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
72	200810_01_AutMiddBb17H2D_72	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
73	200810_01_AutMiddBb17H2D_73	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
74	200810_01_AutMiddBb17H2D_74	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
75	200810_01_AutMiddBb17H2D_75	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
76	200810_01_AutMiddBb17H2D_76	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
77	200810_01_AutMiddBb17H2D_77	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
78	200810_01_AutMiddBb17H2D_78	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
79	200810_01_AutMiddBb17H2D_79	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
80	200810_01_AutMiddBb17H2D_80	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
81	200810_01_AutMiddBb17H2D_81	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
82	200810_01_AutMiddBb17H2D_82	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
83	200810_01_AutMiddBb17H2D_83	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
84	200810_01_AutMiddBb17H2D_84	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
85	200810_01_AutMiddBb17H2D_85	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
86	200810_01_AutMiddBb17H2D_86	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
87	200810_01_AutMiddBb17H2D_87	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
88	200810_01_AutMiddBb17H2D_88	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
89	200810_01_AutMiddBb17H2D_89	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
90	200810_01_AutMiddBb17H2D_90	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
91	200810_01_AutMiddBb17H2D_91	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
92	200810_01_AutMiddBb17H2D_92	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
93	200810_01_AutMiddBb17H2D_93	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
94	200810_01_AutMiddBb17H2D_94	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
95	200810_01_AutMiddBb17H2D_95	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
96	200810_01_AutMiddBb17H2D_96	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
97	200810_01_AutMiddBb17H2D_97	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
98	200810_01_AutMiddBb17H2D_98	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
99	200810_01_AutMiddBb17H2D_99	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N
100	200810_01_AutMiddBb17H2D_100	Glycine betaine (15N)	015N	015N	015N	015N	015N	015N	015N

Python Script

Project Design - Testing

Unit Tests performed to ensure correct data formatting, types in CSV file passed into data processing scripts.

Tests ensure:

- Viable, non-empty CSV is supplied
- Data types within CSV is correct
- CSV contains internal standards to allow for quality control of data



```
"""
This module contains the unit test functions for the nitrogenous fate (nfex)
project

Classes:
- TestKFEX(unittest.TestCase)
"""

import unittest
from nitrogenousfate import process_csv

class TestNFEX(unittest.TestCase):
    """
    TestNFEX(unittest.TestCase): tests the code for data cleaning and wrangling for NFEX.
    """

    def test_smoke_pass(self):
        """
        test_smoke_pass(self): Smoke test (1) verifying function should run without
        crashing or throwing errors if given an appropriate CSV file.
        """
```



Nitrogenous Fate

Susan Garcia, Aaron Lecciones, Brian Roman, Carlyn Schmidgall

CSE583 Software Development for Data Scientists
University of Washington, Seattle

Autumn 2023

Thank you For listening!

