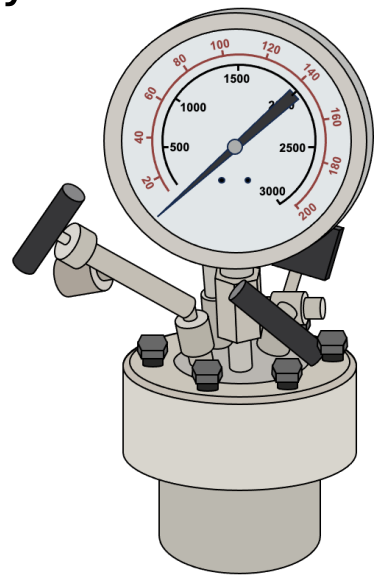
Chroma Quant
Assessment & Extension

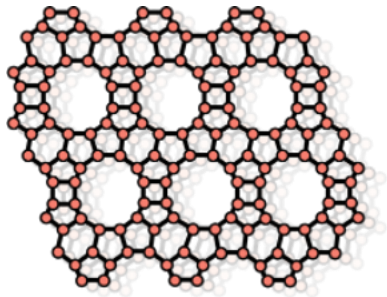
Franklin Guevara • Mariya Hryb • Enisha Sehgal • Tata Serebryany

CHEM E 546 Winter Quarter Project Update

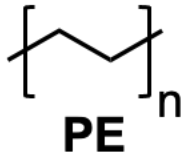
March 18, 2025



Background: Plastic Upcycling



BEA



PE

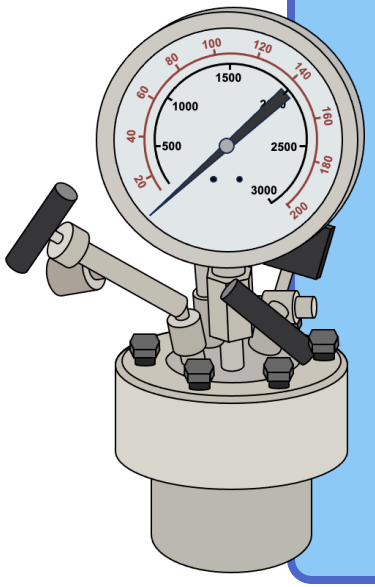
Ru/C

100.1mg Ru/C
95.0mg BEA

1015.6mg PE
(Mw~4k)

72hr, 600rpm, 125psi initial,
~250psi final, N_2

280°C, heat time ~1.5hr

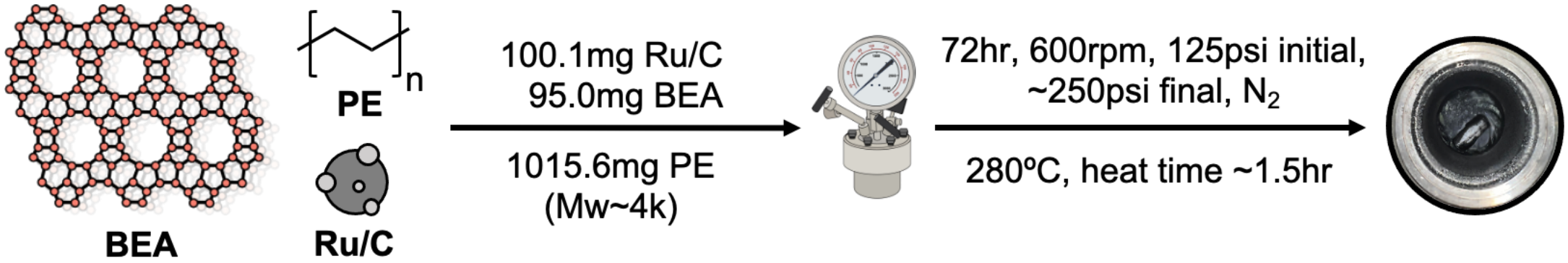


Plastic Polymer
(solid)

- Other gas products
- Other liquid products
- Plastic Monomer (liquid)
- Plastic Polymer (solid)
- Tar (solid)

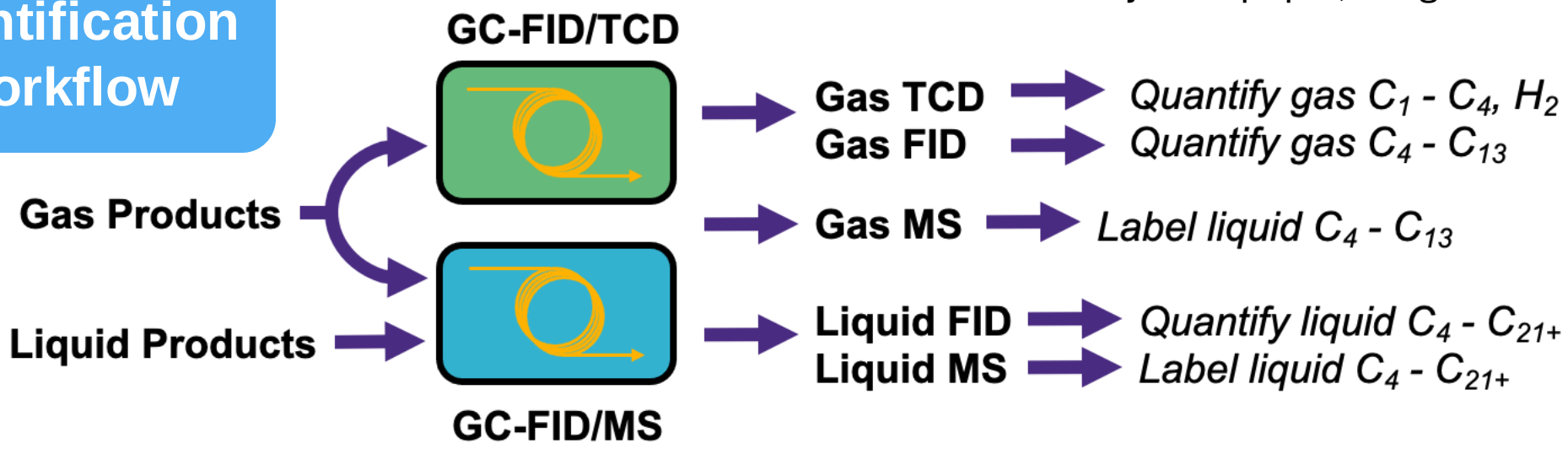
Need a way to analyze the products

Background: Plastic Upcycling



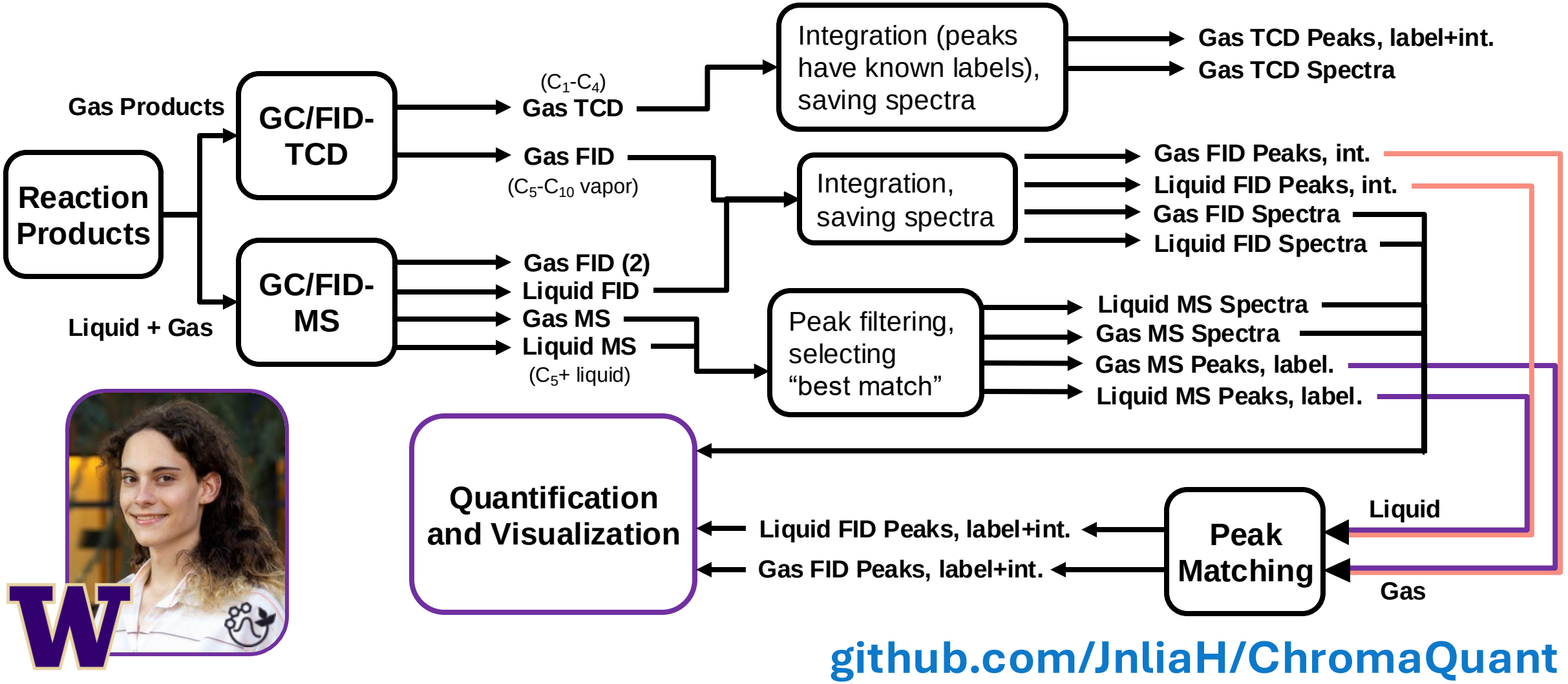
- Gas collected, 10mL CO_2 internal standard
- Solids collected by filter paper, weighed

Quantification Workflow



Hancock and Rorrer (2025) [Manuscript in preparation]

Julia's Creation: Chroma Quant



github.com/JnliaH/ChromaQuant

Data Processing

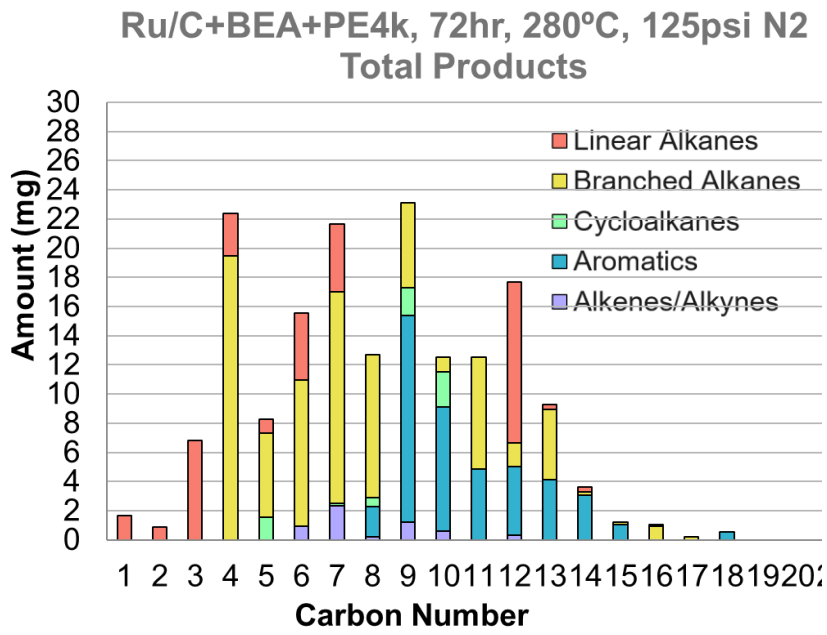
Raw Data from Instruments

- MBPR053_02_GS1_MS_SPEC
- MBPR053_02_GS1_UA_Comp_UPP
- MBPR053_02_GS2_FID_SPEC
- MBPR053_02_GS2_TCD_CSO
- MBPR053_02_GS2_TCD_SPEC
- MBPR053_02_LQ1_FID_CSO
- MBPR053_02_LQ1_FID_SPEC
- MBPR053_02_LQ1_MS_SPEC
- MBPR053_02_LQ1_UA_Comp_UPP

Component Mixture Breakdown

Carbon Number	Aromatics	Linear Alkanes	Branched Alkanes	Cyclo-alkanes	Alkenes /Alkynes	Other
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0.9357	4.9280	1.4413	0.0955	0.0048
6	0	0.2258	1.6310	0.0587	0.8955	0.0002
7	0	0	0.6621	0.0165	0.0967	0
8	0.0015	0	0.1285	0	0.0117	0
9	0.0001	0	0.0105	0.0001	0.0012	0.0038
10	0	0	0.0002	0.0000	0	0
11	0	0	0	0	0	0
12	0	0	0	0	0	0
13	0	0	0	0	0	0
14	0	0	0	0	0	0
15	0	0	0	0	0	0.0007

Stacked Bar Chart of Components by Carbon Number



Pre-processed with Agilent MassHunter Quantitative Analysis Software

Option: Manual or Chroma Quant

How consistent is it? Only some anecdotes

Liquids Analysis

Maddie's Manual Quantification

ret	name	formula	Area	Mass of liq (mg)
2.4390576	Isobutane	C4H10	0.449115	0.378199775
6.5737635	Cyclohexane, 1,4-dimethyl-	C8H16	1.23536	1.040296748
6.8663	Cyclopentane, 1-ethyl-1-methyl-	C8H16	0.57493	0.484148596

Julia's Automatic Quantification

FID RT	FID Area	MS RT	Compound Name	Formula	m_i
2.443	0.4038	2.439057565	Isobutane	C4H10	0.904900608
6.341	0.2098	6.332681415	No Match		
6.433	2.5295	6.430323764	Heptane, 3-methyl-	C8H18	17.51241462
6.576	0.8741	6.573763503	Cyclohexane, 1,4-dimethyl-	C8H16	1.624257867
6.809	0.3706	6.807704841	1,5-Heptadien-4-one, 3,3,6-trimethyl-	C10H16O	0.688651145

Build a software to compare the accuracy of ChromaQuant!

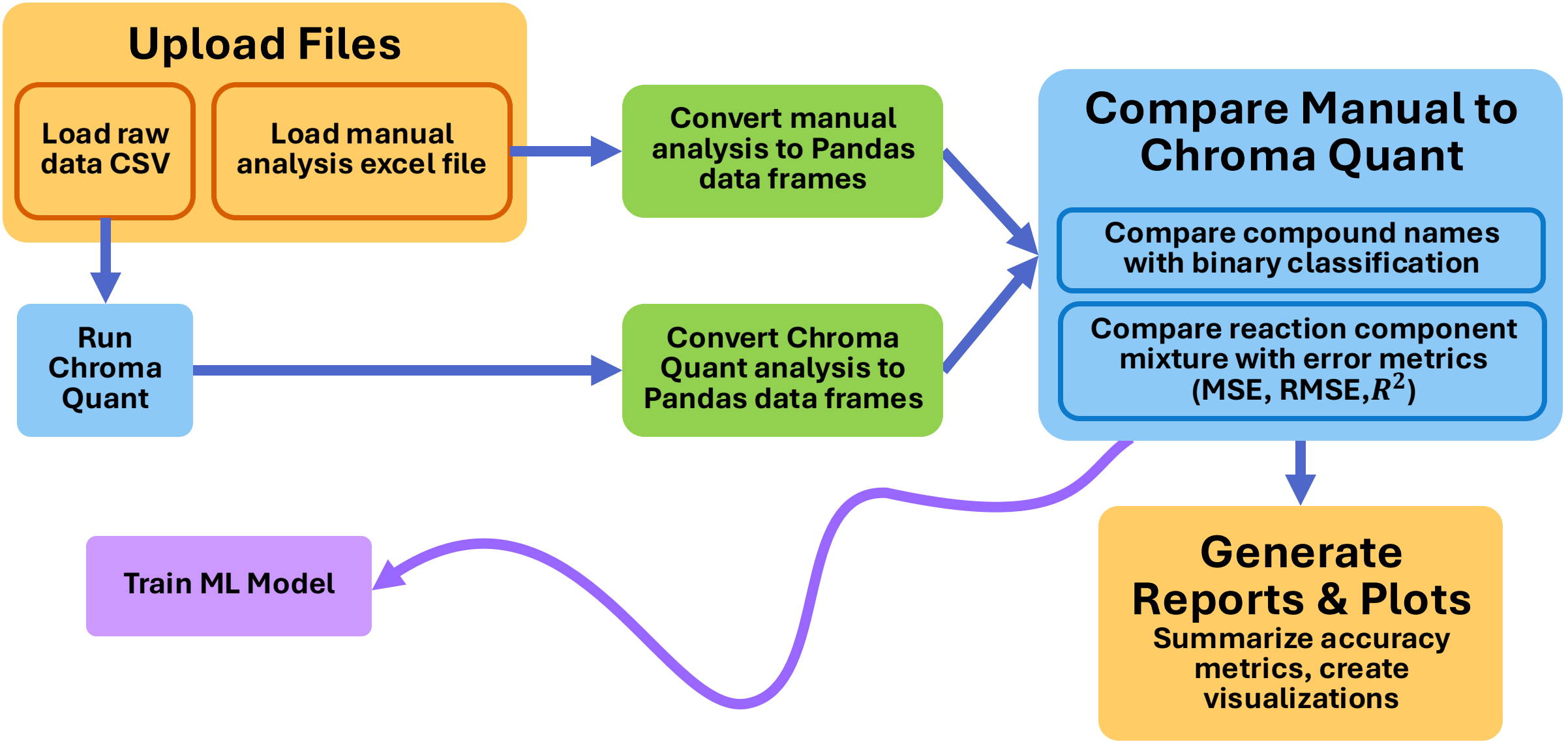
(and add an ML model extension to ideally improve performance)

Identification was good, but **integration results seemed inconsistent**

User Stories for Chroma Quant Extension

User Category	Goals	Needs & Wants	Skill Level
CQ developers (Julia)	Compare the output of manual and automatic analyses	A report on the accuracy of the software.	She knows everything about Chroma Quant because she made it
CQ internal users (Rorrer lab)	Use the assessment tool to confirm by-hand calculations	A way to input additional data into assessment tool	Comfortable with GC/MS FID/TCD. Okay with using CQ
CQ external users (other researchers)	Want to see proof that CQ works before they use it in their lab	Easy-to-read figures and graphs assessing the CQ accuracy	Comfortable with GC/MS FID/TCD. No python, UI only.

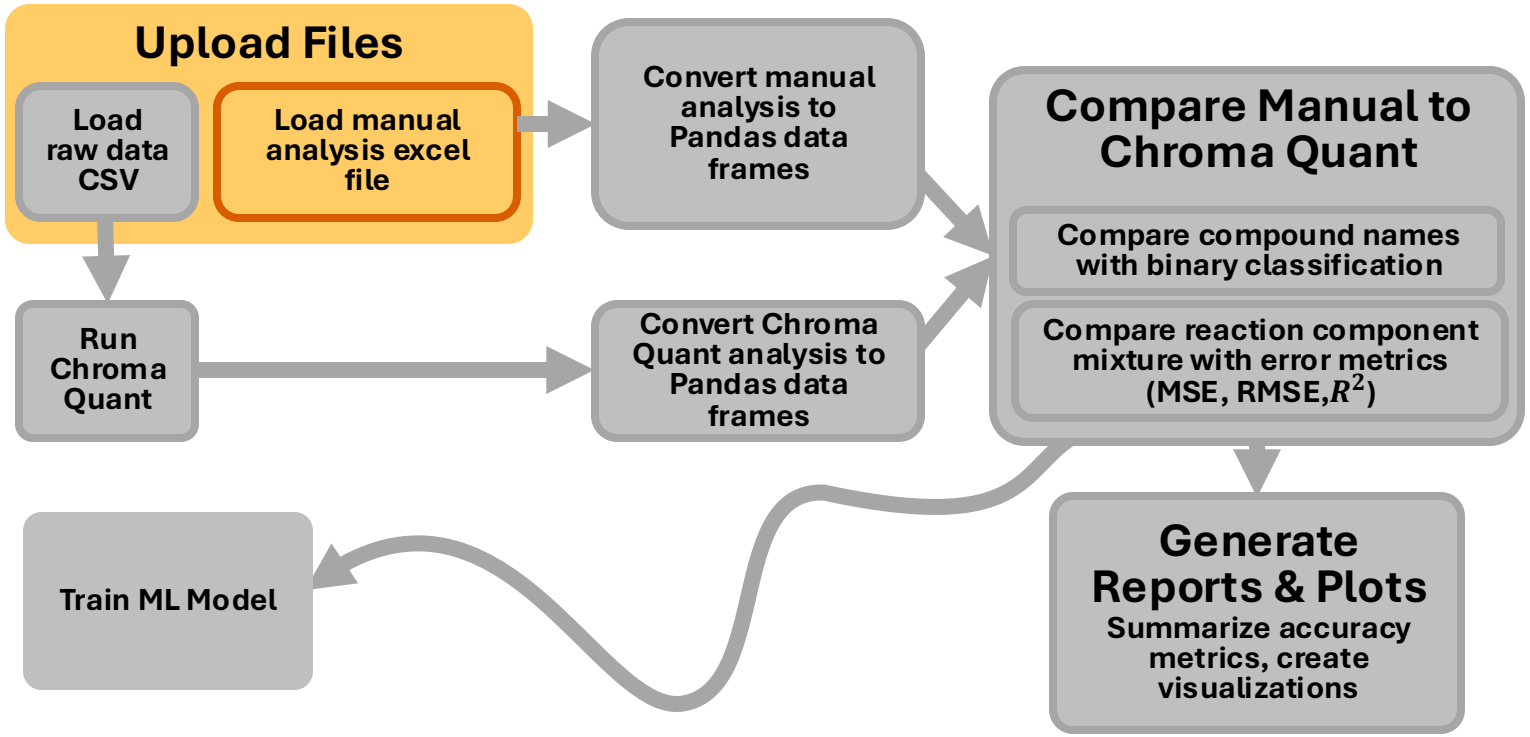
Assessment Tool Flowchart



Closer Look: Components & Their Tests

1. CSV Manual Loader Test

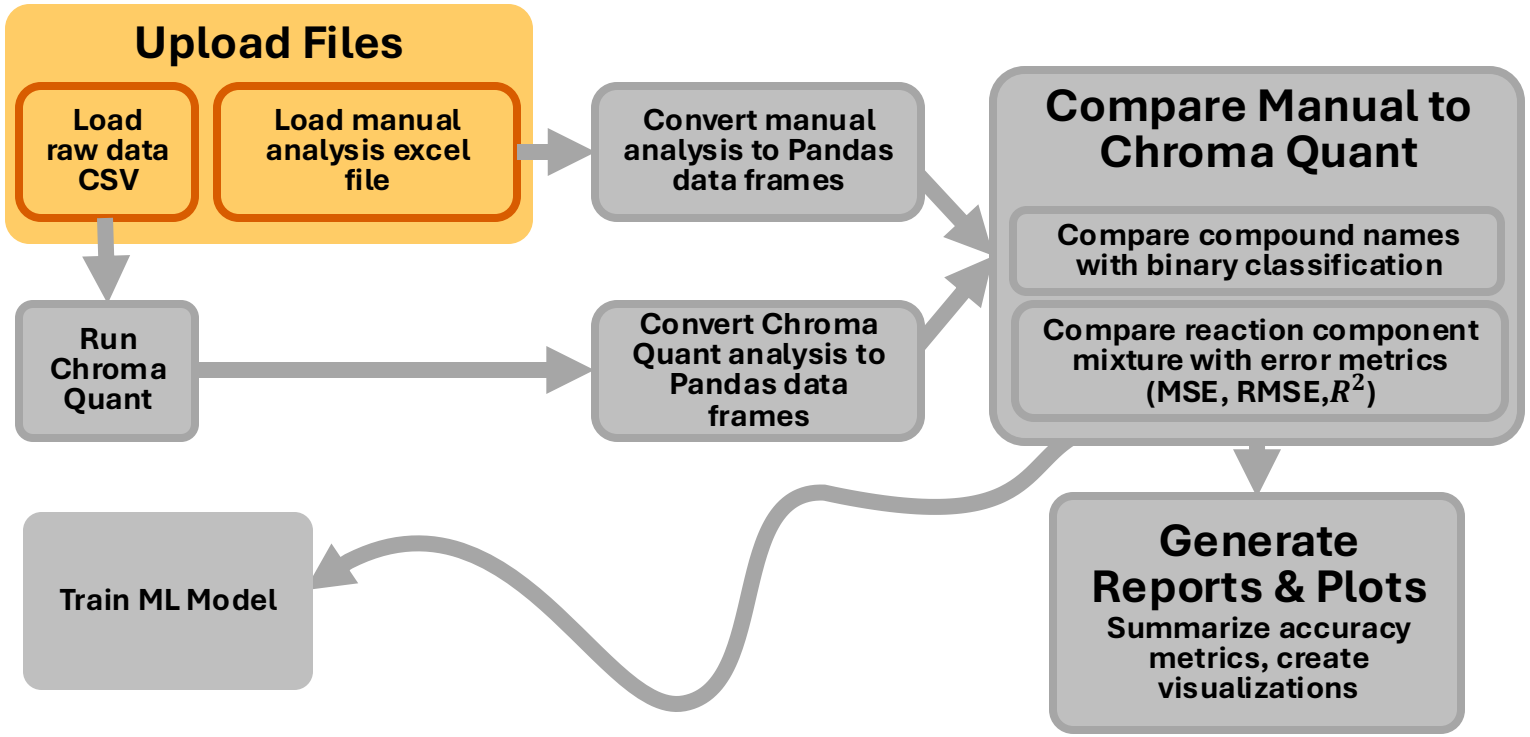
- Simulates a manual analysis CSV file
- Tests that it has correct columns
- Raises error if no file is provided



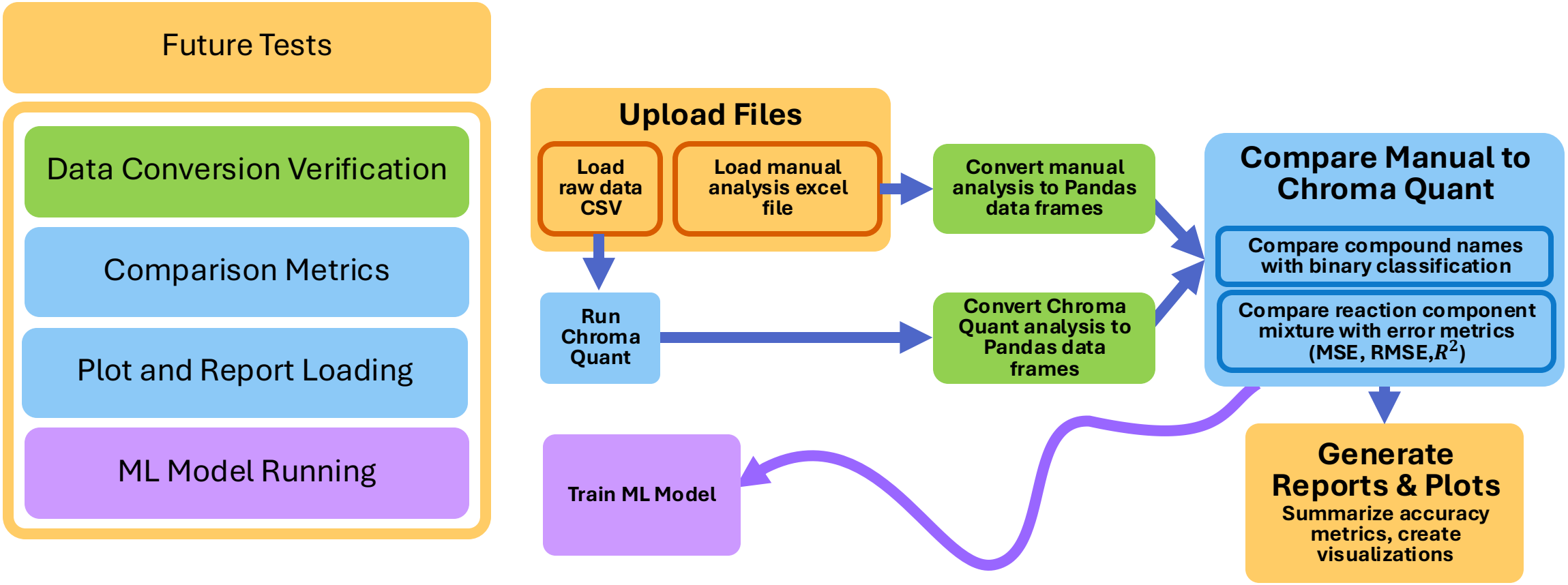
Closer Look: Components & Their Tests

2. ChromaQuant CSV Formatting Test

- MBPR053_02_GS1_MS_SPEC
- MBPR053_02_GS1_UA_Comp_UPP
- MBPR053_02_GS2_FID_SPEC
- MBPR053_02_GS2_TCD_CSO
- MBPR053_02_GS2_TCD_SPEC
- MBPR053_02_LQ1_FID_CSO
- MBPR053_02_LQ1_FID_SPEC
- MBPR053_02_LQ1_MS_SPEC
- MBPR053_02_LQ1_UA_Comp_UPP



Closer Look: Components & Their Tests



Progress so Far...

Code | Blame | 21 lines (19 loc) · 350 Bytes

```yaml
name: chromaquantenv
channels:
  - conda-forge
dependencies:
  - python=3.10
  - pip
  - pandas>=2.2.2
  - molmass>=2024.5.24
  - numpy>=1.26.4
  - scipy>=1.13.1
  - pillow>=10.3.0
  - plotly>=5.22.0
  - pubchempy>=1.0.4
  - rdkit>=2024.3.5
  - XlsxWriter>=3.2.0
  - openpyxl>=3.1.5

  - pip:
      - chemformula~=1.3.1
      - ttkthemes~=3.2.2
```

An environment for users to create:

```
conda env create -f environment.yml -n chromaquantenv
```

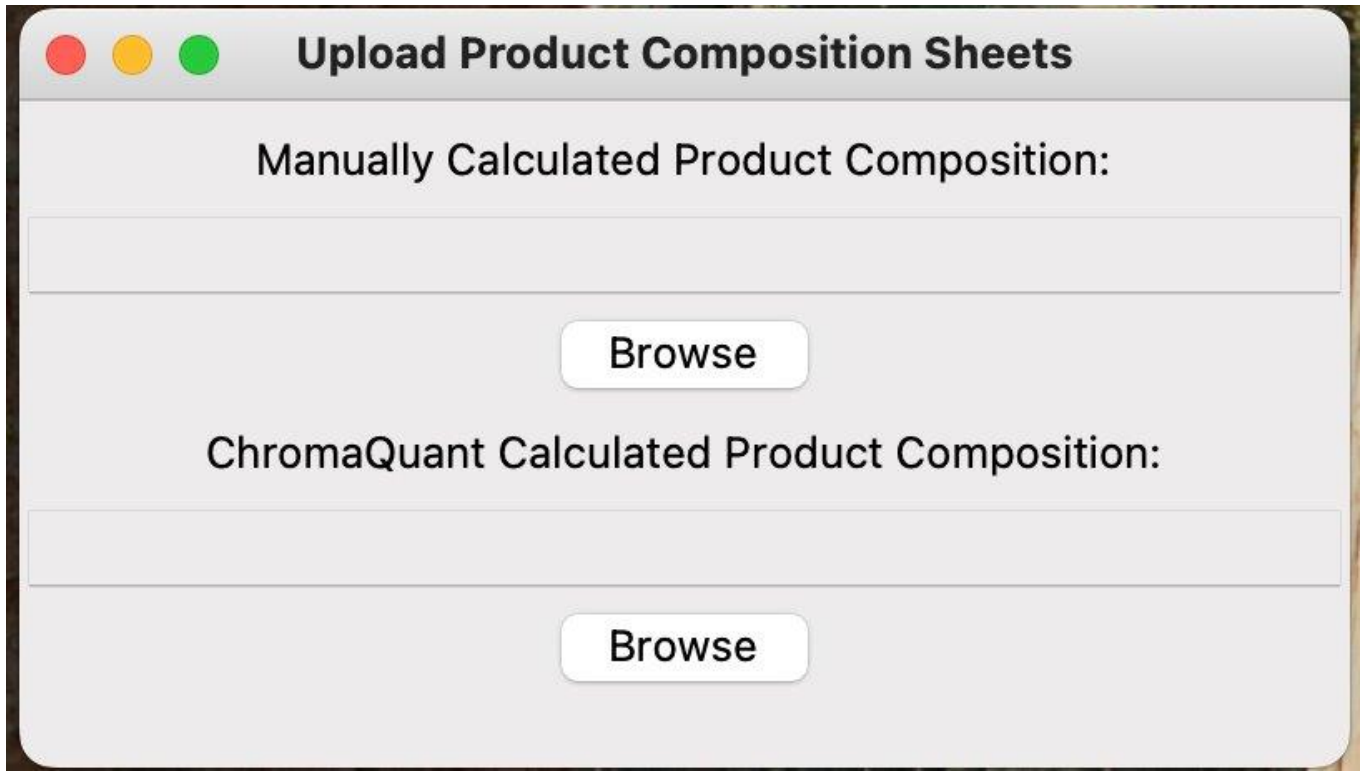
ChromaQuantExtension is pip Installable

```
pip install -e./ChromaQuantExtension
```

```
conda list
```

```
chromaquantextension      1.0.0                    pypi_0    pypi
```

A Preliminary Interface for the User

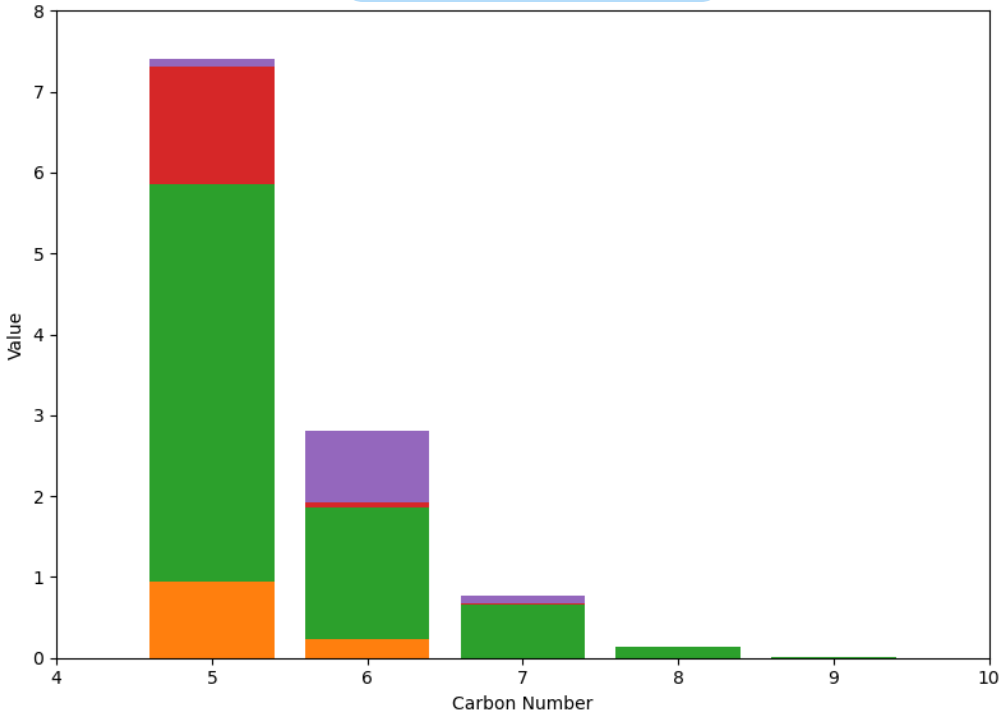


User uploads their manually calculated csv

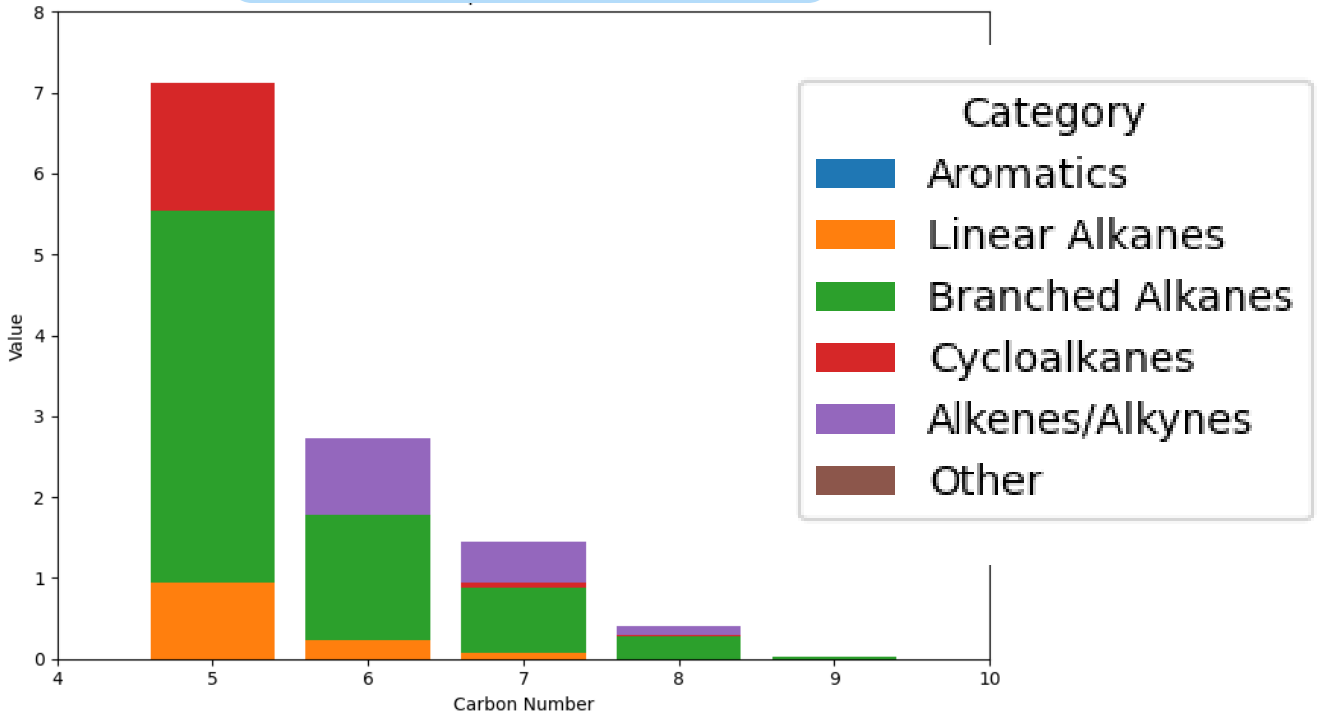
User uploads ChromaQuant calculated csv

Preliminary Results: Liquid Products of 1 rxn

Manual

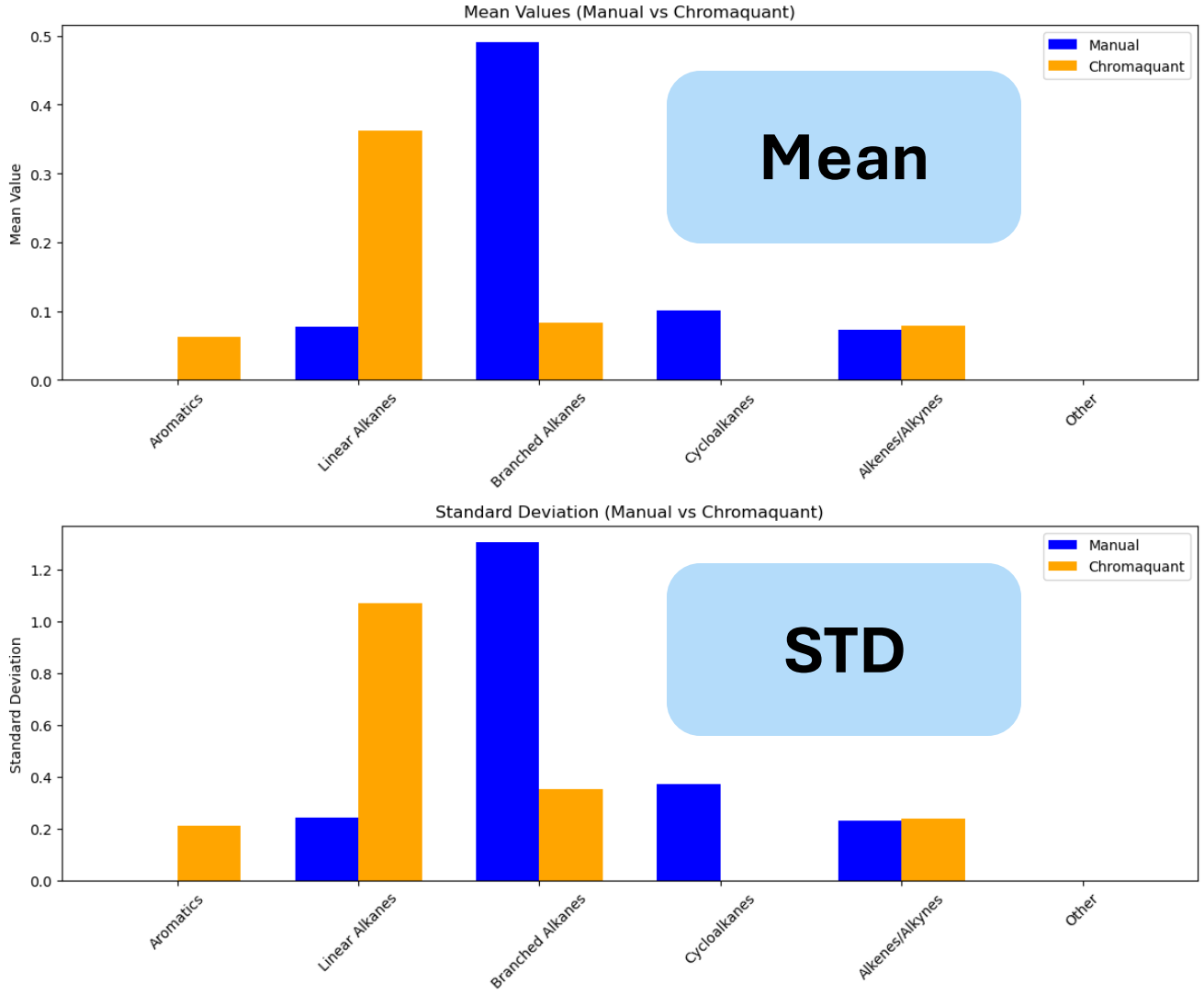


Chroma Quant



Not insignificant differences in each category

Preliminary Results: Liquid Products of 1 rxn



Questions to Answer

- How to compare MANY data sets?
- What metrics to assess?
- What counts as “success”?

Future Work

Better File Management

Big Data Analysis

Implement ML

Conda Package

	Quarter 1	Quarter 2									
		March 31	April 7	April 14	April 21	April 28	May 5	May 12	May 19	May 26	June 2
Logo											
Git Repo											
Upload Files											
Run Chroma Quant											
Convert CQ to df											
Convert manual to df											
Compare (Math)											
ML extension											
Tests											
Conda package											

What We Wish We Wknew Wearlier

- Resolving merge conflicts
  - Git prune

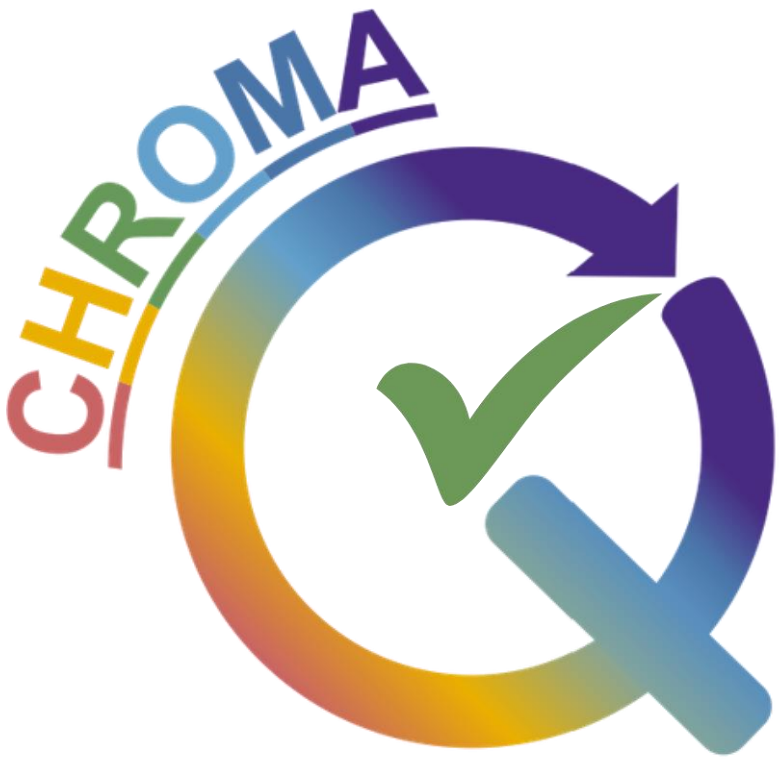
- Unexpected data formatting

- Dependency issues
  - Conda forge vs. pip install

python™

ANACONDA
Powered by Continuum Analytics

Thank you for listening!

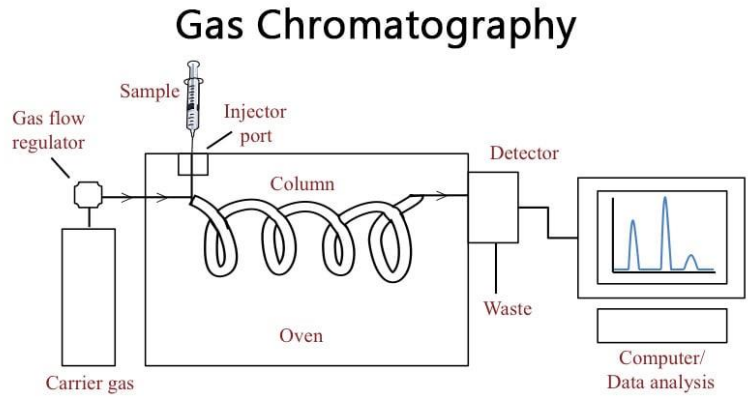


Full Components Chart

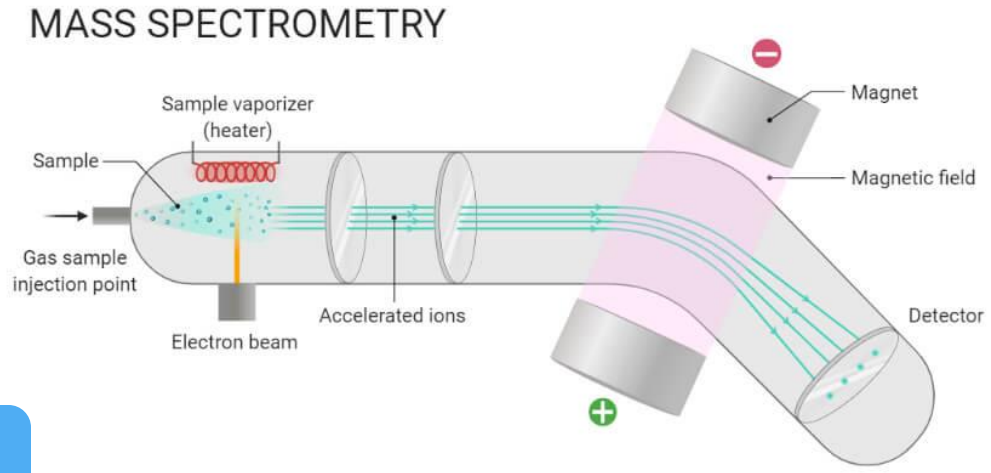
Component Name	What It Does	Inputs (Type Info)	Outputs (Type Info)	How It Uses Other Components	Potential Side Effects
Upload Files	Loads manual data files and CQ input files, converting them to Pandas DataFrames	CSV/Excel (strings, ints, floats)	Pandas DataFrames (strings, ints, floats)	-	File format errors (unsupported file types, encoding issues), corrupt/missing data (empty or incomplete files)
Run CQ Analysis	Runs ChromaQuant on manual data to generate a CQ-processed dataset	CSV of manual data (strings, ints, floats)	CSV of CQ output (strings, ints, floats)	Uses Upload Files	CQ execution failures (software crashes, incorrect input format), unexpected output format
Convert Data to Pandas	Converts CQ and manual output files to Pandas Data Frames	CSV file paths	Pandas DataFrames (strings, ints, floats)	Uses Run CQ Analysis	Parsing errors (wrong delimiters, inconsistent column headers), NaN values due to missing data
Preprocess Data	Cleans and normalizes CQ and manual data for easier comparison	Pandas DataFrames (strings, ints, floats)	Cleaned Pandas DataFrames	Uses Convert Data to Pandas	Data loss (incorrect handling of missing values), data misalignment (mismatched indices or column names)
Compare Compound Names	Performs binary classification to check CQ compound name accuracy	Pandas DataFrames (manual & CQ)	Pandas DataFrame with binary classification (T/F or 0/1)	Uses Preprocess Data	False positives/negatives (misclassifications due to inconsistencies in naming conventions)
Compare Reaction Component Mixture	Computes error metrics (MSE, RMSE, R^2) for reaction component mixture analysis	Pandas DataFrames (manual & CQ)	Pandas DataFrame with error values (floats)	Uses Preprocess Data	Skewed error metrics (outliers heavily affecting MSE), divergent scales (unit mismatch)
Generate Reports & Plots	Summarizes accuracy metrics, creates visualizations	Pandas DataFrames from comparison components	CSV report & plots	Uses Compare Compound Names & Compare Reaction Component Mixture	Plot rendering issues (if input data is missing or poorly structured), incorrect summary statistics due to bad preprocessing
Train ML Model	Trains a Random Forest model if CQ accuracy > 95%	CSV from reports (strings, ints, floats)	ML model for predicting reaction component mixture	Uses Generate Reports & Plots	Overfitting (if dataset is too small or unbalanced), poor generalization (if CQ accuracy is unreliable)

Background: Analysis Instruments

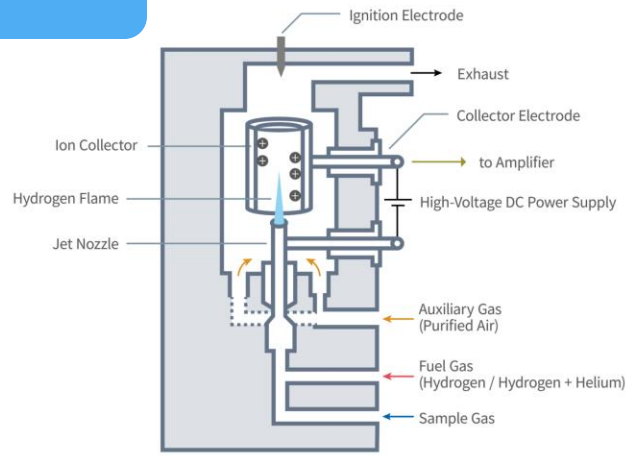
GC



MS



FID



TCD