

AI as a Service (AlaaS) in the RWD/RWE Space

PHUSE US Connect 2021, 14Jun2021 – 18Jun2021

Aaron Clark

Senior Data Scientist, Data Sciences, Biogen

Himanshu Pandya,

Senior Principal Data Scientist, Data Sciences, Biogen

Bob Engle,

Head of Data Sciences, Biogen

DISCLAIMER

The scope of this presentation is to present the opinions and suggestions of the authors. The interpretations of standards and procedures contained in this paper are those of the authors and may not be necessarily correct. Any views and recommendations stated within this paper are those of the authors, and they do not represent the positions of their employer.

Agenda

- ❑ Introduction
- ❑ What is AI as a Service (AlaaS)?
 - It's significance and how it's shaping the life science industry
- ❑ Introduce two AlaaS products – TidyRename, QuickClusters (R Shiny Apps)
- ❑ Demos!
- ❑ Conclusions, lessons learned (so far...)

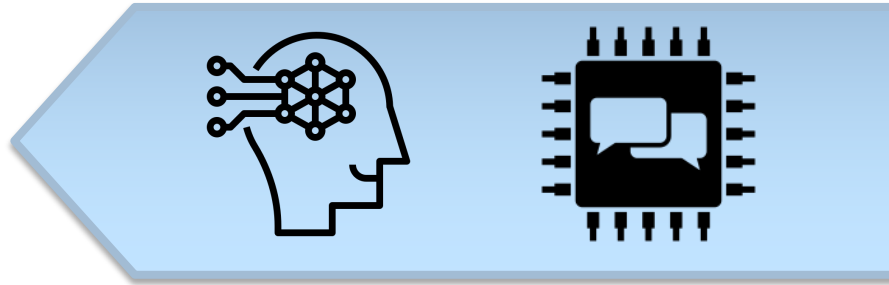
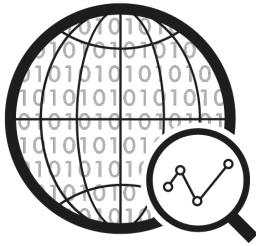
Introduction

- ❑ The Data Science team at Biogen recently launched “Artificial Intelligence as a Service” (**AlaaS**) concepts in the RWD/RWE space using the R statistical software.
- ❑ AlaaS allows individuals/teams to experiment with AI for various project purposes with any level of expertise and without any major initial setup/prep work.



Introduction (cont'd)

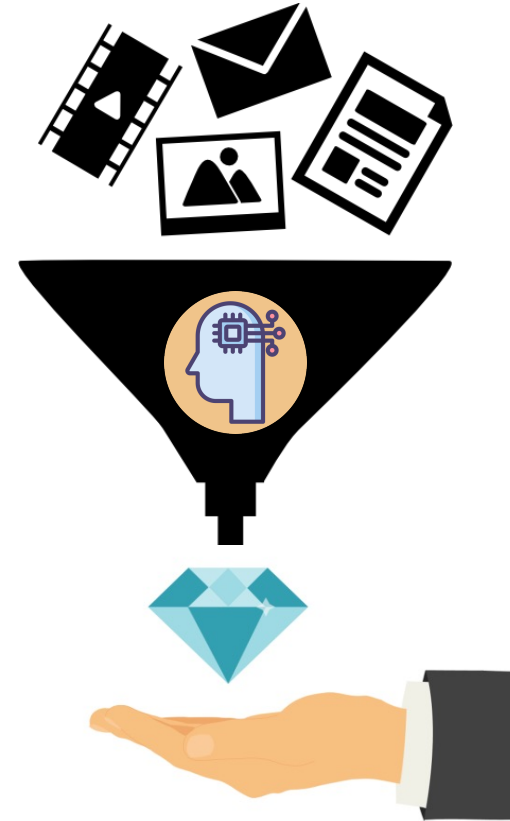
- ❑ Two elements of AI that stand out as being particularly beneficial for handling RWD - Machine Learning (ML) and Natural Language Processing (NLP).



- ❑ We'll address:
 - salient features of the AI initiatives (ML, NLP) in the RWD/RWE space
 - the overall design/implementation process
 - challenges faced / lessons learned

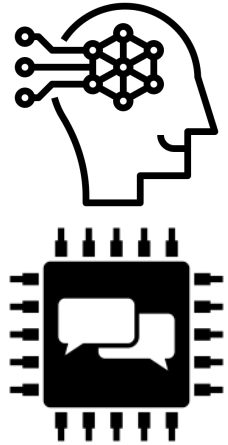
Significance of AI in the Life Science Industry

- ❑ Amounts of real-world data (RWD) are increasing from a variety of sources - electronic health records, medical imaging, insurance records, wearables and health apps, social media, clinical trials, genomic sequencing etc.
- ❑ A key differentiator for biopharma companies is the extent to which they can generate insights and evidence from multiple data sources, which makes digital transformation a strategic imperative and AI has the full potential to accelerate.

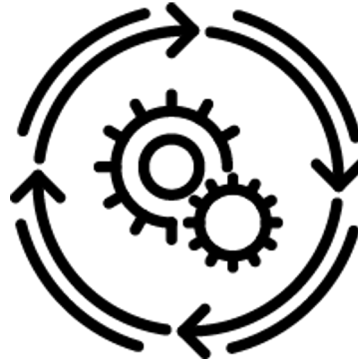
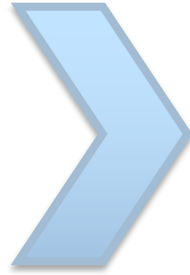


Significance of AI in the Life Science Industry (cont'd)

- ❑ Ultimately, AI such as machine learning (ML) and natural language processing (NLP) can automate numerous processes such as monitoring activities, generating reports, managing queries and others that are tedious, error prone, and consume highly skilled resources.



ML & NLP



Automation



Reduce Waste

Two AI initiatives (R Shiny Apps) launched in the RWD/RWE space.

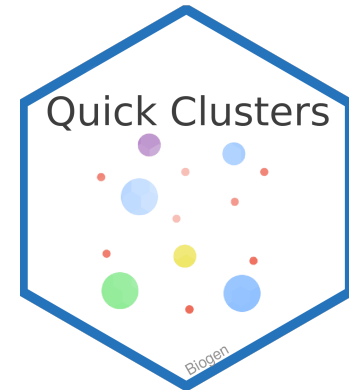
❑ TidyRename

Reduces project work redundancy across multiple studies by automating the creation of variable names. Concepts can be expanded to CSR activities i.e., SDTM variables mapping process etc.



❑ QuickClusters

To support the Public/External data-based research activities using unsupervised algorithms. Clusters/PCA methods leveraged to generate valuable insights on the prodigious incoming RWD and harness it's potential to support data driven decisions.



Application's prominent features:

- ❑ App helps users generate submission compliant variable attributes
 - i.e. ≤ 8 characters for variable names, ≤ 40 characters for labels
- ❑ User friendly point-and-click functionality. Upon uploading user data, the variable abbreviation algorithm quickly produces renamed suggestions and a frequency chart of the transformation methods applied to the raw data
- ❑ Users has an option to overwrite suggestions and upload their own dictionary of preferred terms. Users can also append/alter the dictionary per business needs
- ❑ Ensures uniformity among users and eliminates wasted time manually creating easily-readable variable names for data sets with many columns.

TidyRename (Cont.)

To summarize...

	subject id	Weight_1	1b. Eyes	Resting_5_1	Rigidity_7_2	AgeDiagnosedPD_3	Standing_Diastolic_7b
0	1	97	1	0	1	61	156
0	1	97	1	0	1	61	156
0	1	97	1	0	1	61	156
0	1	97	1	0	1	61	156
0	2	65	0	1	NA	64	70
0	3	54	0	0	1	57	76
0	3	54	0	0	1	57	76
0	3	54	0	0	1	57	76

Input Data
File



Output
Data File

V1	SUBJID	WEIGHT_1	1BEYES	REST_5_1	RIGID72	AGDGNPD3	STNDDS7B
0	1	97	1	0	1	61	156
0	1	97	1	0	1	61	156
0	1	97	1	0	1	61	156
0	1	97	1	0	1	61	156
0	2	65	0	1	NA	64	70
0	3	54	0	0	1	57	76
0	3	54	0	0	1	57	76
0	3	54	0	0	1	57	76

Step 1 - Upload/Preview your data



tidyRename

Upload Data

Rename Suggestions

Methods & Stats

Alter Dictionary

Data Upload/Preview



View example data

Use example data

Use Random data

Select an Excel or CSV file

Browse...

my_data.csv

☐ More Upload Options

Upload complete

Preview:

Show entries

Search:

subject	year	visit	1a.	1a.	1a.	1a.	1b.	1b.	1b.	1b.	1c.
id	number	date	Skin	Skin_Desc	Skin_Active/resolved	Skin_Date diagnosis	Eyes	Eyes_Desc	Eyes_Active/resolved	Eyes_Date diagnosis	ENT_Desc

Step 2- Variable name creation



tidyRename

Upload Data

Rename Suggestions

Methods & Stats

Alter Dictionary

Rename Suggestions



Customize Suggestion Engine



Export Suggestions

Export Renamed Data

Warning: could not render a unique suggestion <= 8 characters for cols: 458, 491.

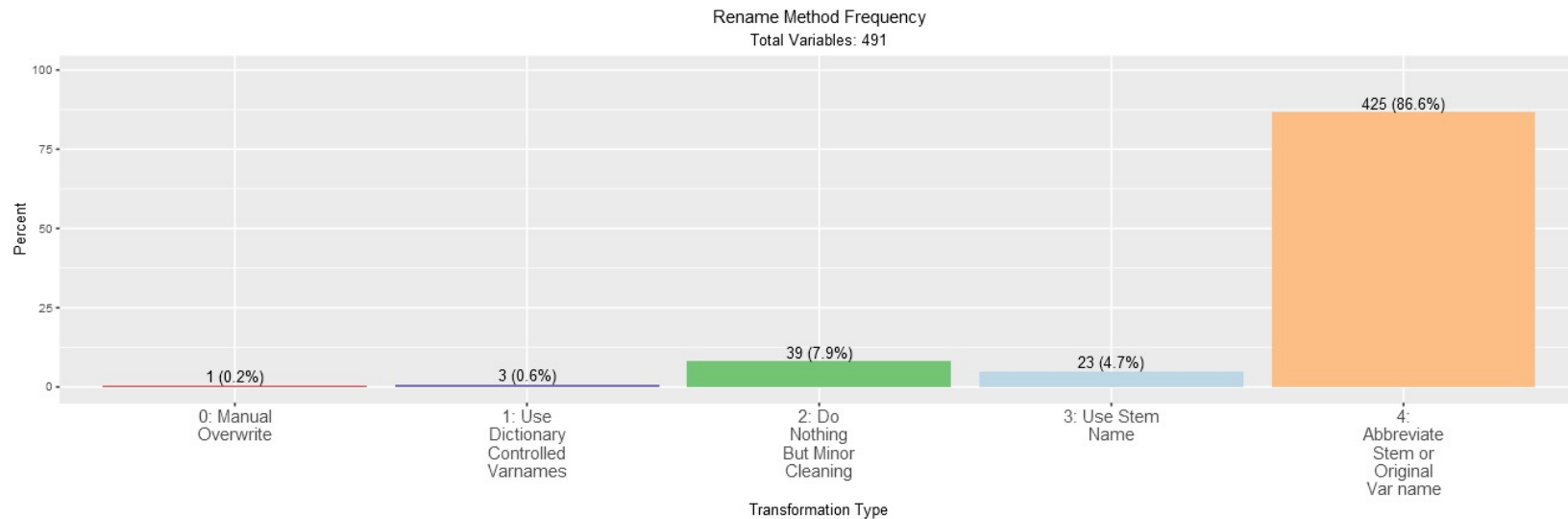
Please overwrite duplicates.

	ORIGINAL_VARNAME	COL_POS	SUGG_VARNAME	OVERWRITE	OVERWRITE_REASON	LABEL
1	subject id	1	SUBJID			Subject Id
2	year number	2	YEARNMBR			year number
3	visit date	3	VISITDAT			visit date
4	1a. Skin	4	_1A_SKIN			1a. Skin
5	1a. Skin_Desc	5	_1ASKNDS			1a. Skin Desc
6	1a. Skin_Active/resolved	6	_1ASKNAR			1a. Skin Active/resolved
7	1a. Skin_Date diagnosis	7	_1ASKNDD			1a. Skin Date diagnosis
8	1b. Eyes	8	_1BEYES			1b. Eyes
9	1b. Eves Desc	9	1BEYDSC			1b. Eves Desc

Step 3 - View Rename Method Frequencies

tidyRename Upload Data Rename Suggestions **Methods & Stats** Alter Dictionary

Methods & Stats



Step 4 – Alter Dictionary

[tidyRename](#)[Upload Data](#)[Rename Suggestions](#)[Methods & Stats](#)[Alter Dictionary](#)

Alter Dictionary

[Manually Edit](#)[Upload my Own](#)[Toggle Off / On](#)

Manually Alter the Dictionary:

Add a term:

Original Var-name

New Var-name

New Var-name Label

TidyRename - Demo



Application's prominent features:

- ❑ User-friendly app to perform K-Means clustering algorithm whilst managing multiple combinations of the following options:
 - number of desired clusters
 - response variables with special handling for BDS data
 - standardized scores vs non-standardized
- ❑ Prior to feeding the algorithm data, users have choice to filter the uploaded data
- ❑ Output results are displayed in both plot and tabular format, which can be exported for additional analysis requirements

Step 1 (Upload/Preview the data)



Quick Cluster

Data

Clustering

Data Upload/Preview

Use 'Test Data'

Data upload

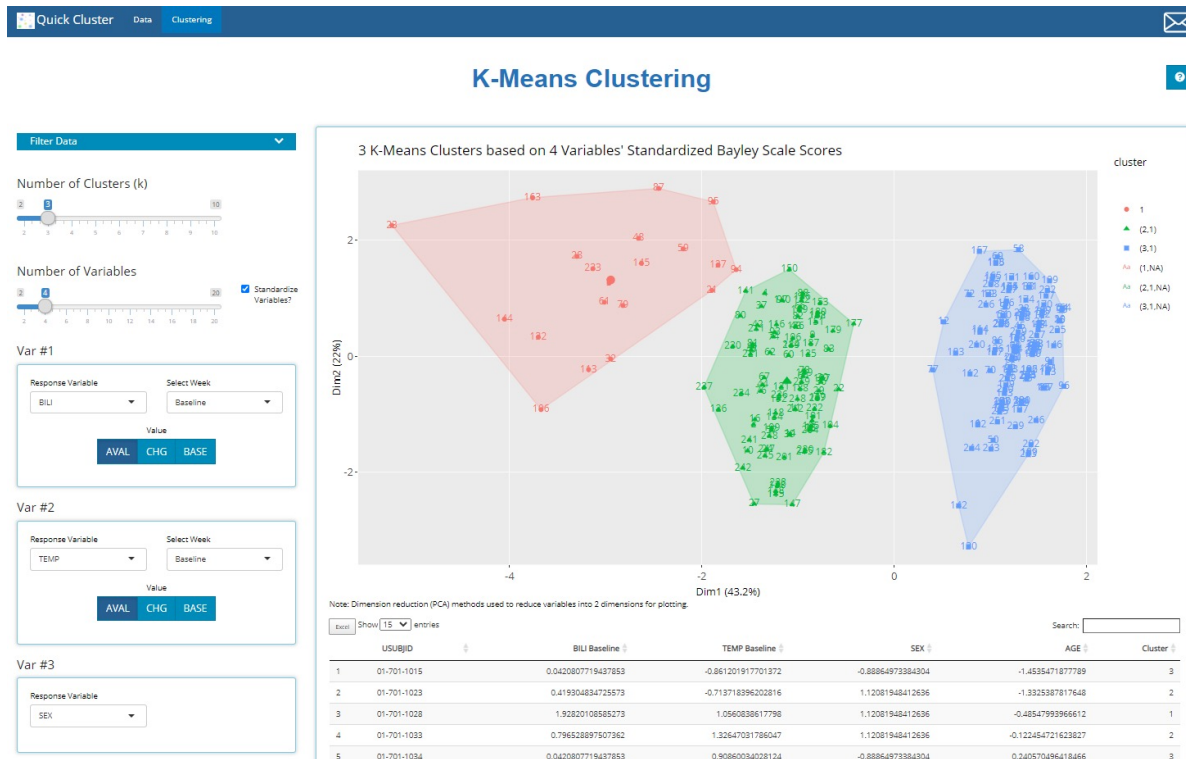


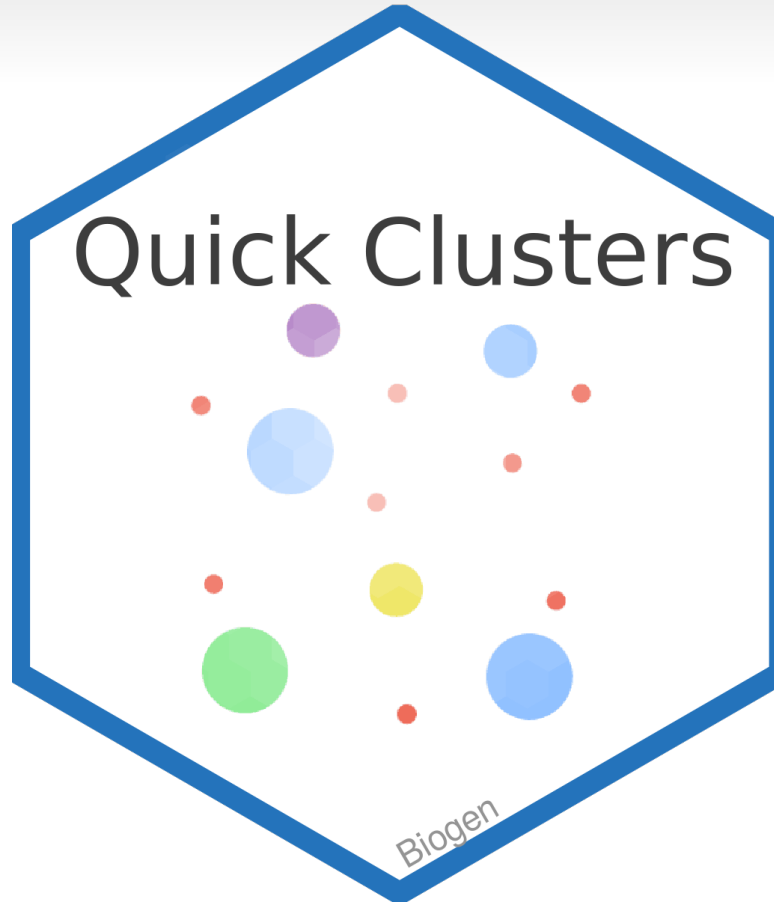
Upload csv or sas7bdat files

Browse...

No file selected

Step 2 - Data filter, selection of cluster characteristics and cluster output





Conclusion

- ❑ The main objective of developing these Apps is to let users explore AI concepts with any level of expertise and without any major prep work.
- ❑ Unsupervised algorithms like clustering techniques are great mechanisms to understand cohort distribution, especially when classified on certain patient characteristics like disease or duration characteristics, demographics, etc. It provides valuable insights to support scientific data driven decisions.
- ❑ NLP can automate numerous processes to reduce redundancies, costs, and improve overall quality by eliminating/reducing human interactions/errors.
- ❑ RWD offers the best fodder for AI methods, however the methods can be expanded to the RCT data world as well. With additional enhancements, the tidyRename app can be expanded to RCT/CSR activities; i.e. SDTM variables mapping process, etc.
- ❑ These are the first version of the applications; we'll keep updating them with new enhancements; i.e. adding supervised algorithms like neural networks, decision trees, linear regression, etc.

Contact Information

Your comments and questions are valued and encouraged. Authors can be contacted at:

Aaron Clark,
Senior Data Scientist,
Biogen, Cambridge, MA
Aaron.Clark@Biogen.com

Himanshu Pandya,
Senior Principal Data Scientist,
Data Sciences,
Biogen, RTP, NC
Himanshu.Pandya@Biogen.com

Bob Engle,
Head of Data Sciences,
Biogen, Cambridge, MA
Bob.Engle@Biogen.com