POLYNARY

Data Visualization Software

User Guide

https://polynaryselfservice.azurewebsites.net/

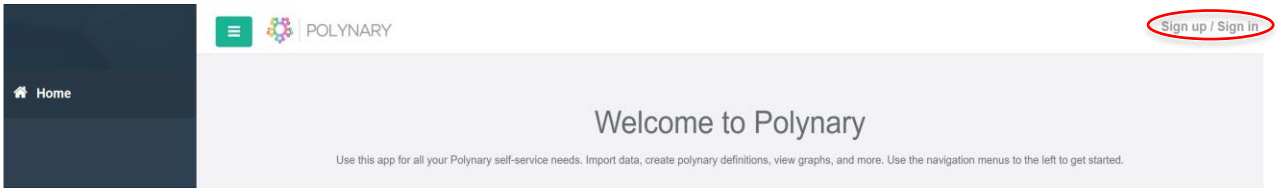
Table of Contents

Creating a Polynary account

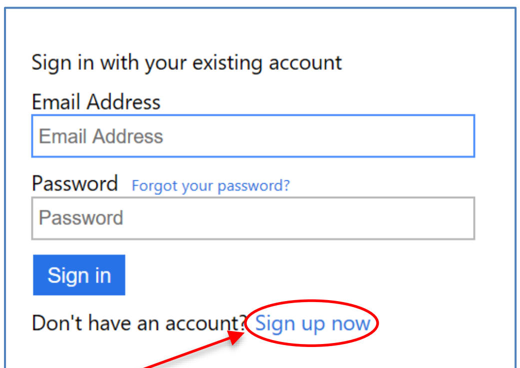
To create a Polynary account, go to Polynary Service site:

https://polynaryselfservice.azurewebsites.net/

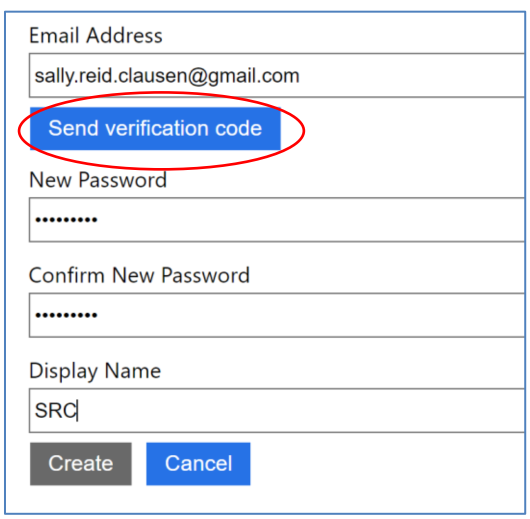
Click on "Sign up / Sign in " in the upper right corner of the Welcome Screen.



You'll be directed to a Sign-in screen:

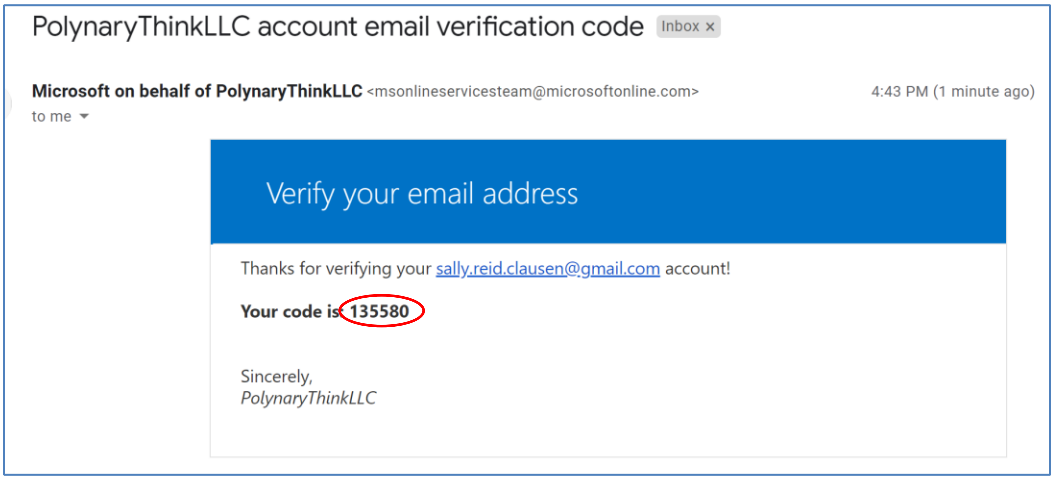


Click on Sign up now link. Complete the sign-up form and click on ***Send verification code*** button.

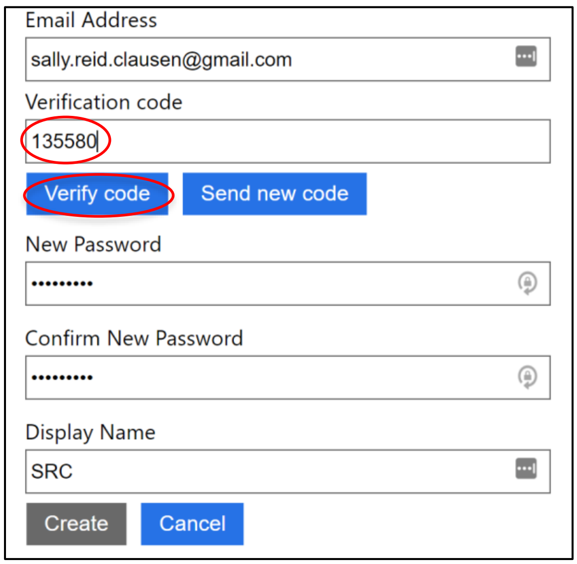


An email is generated with your 6 digit code.

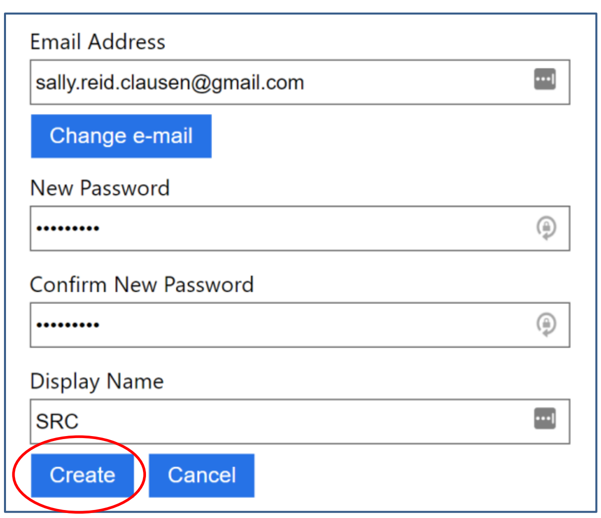
Go to your email and copy verification code



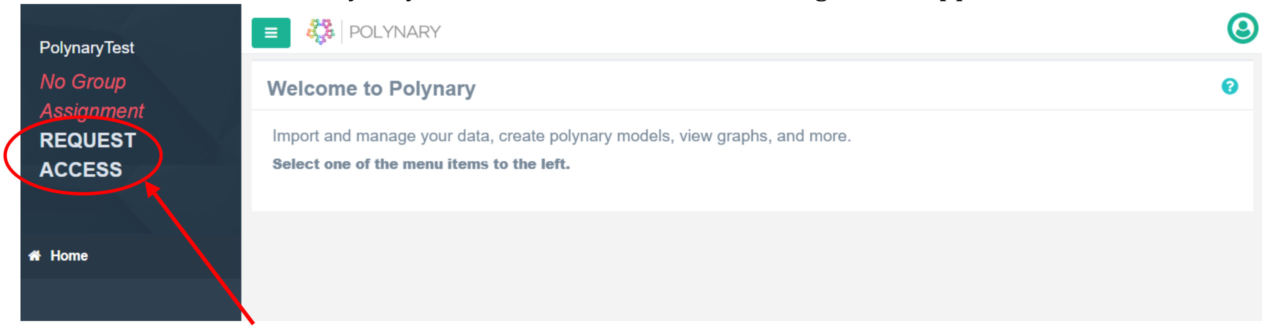
Paste the code into Verification code field, and click on ***Verify code*** button



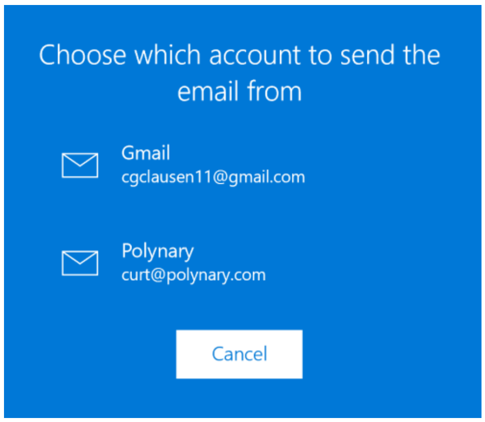
Click the *Create* button, which has now become active (changed from gray to blue).



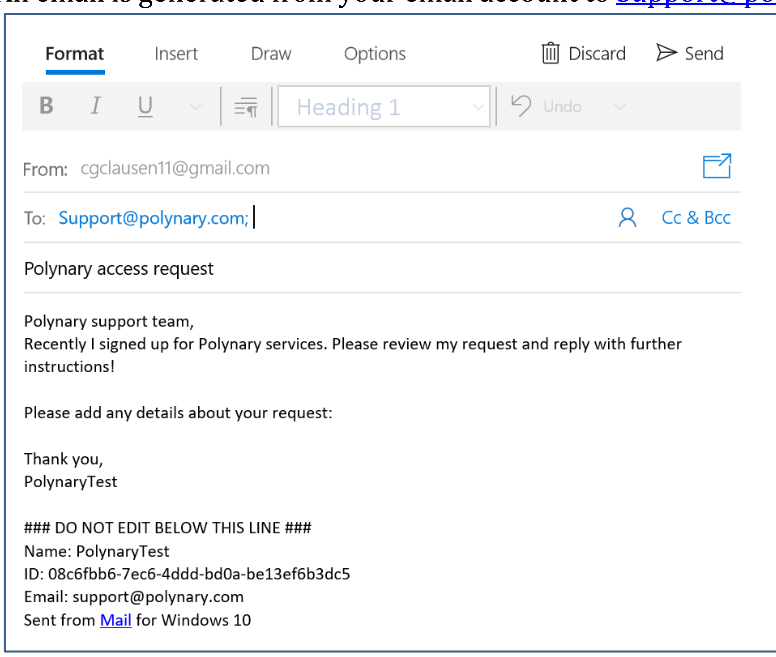
You will be directed to Polynary Service Site where the following screen appears



Click on **Request Access.** If you have multiple email clients, you're prompted to select the account from which to send the request.



An email is generated from your email account to Support@polynary.com.



You may modify the body of the email to include additional information, but *PLEASE DO NOT MODIFY* the area below the" ###DO NOT EDIT BELOW THIS LINE###" warning. This information is needed for us to complete your account set up.

If you are the first person in your organization to request Polynary access, please ***provide a team name***. We will use that to create a *Team Data Library* for your group. Subsequent members who request access should also mention this team name so we can grant them access to the same Team Data Library.

Click *Send.*

You'll receive a *Welcome to Polynary* e-mail notification when your account setup is within 24 hours.

Preparing Polynary Data Files

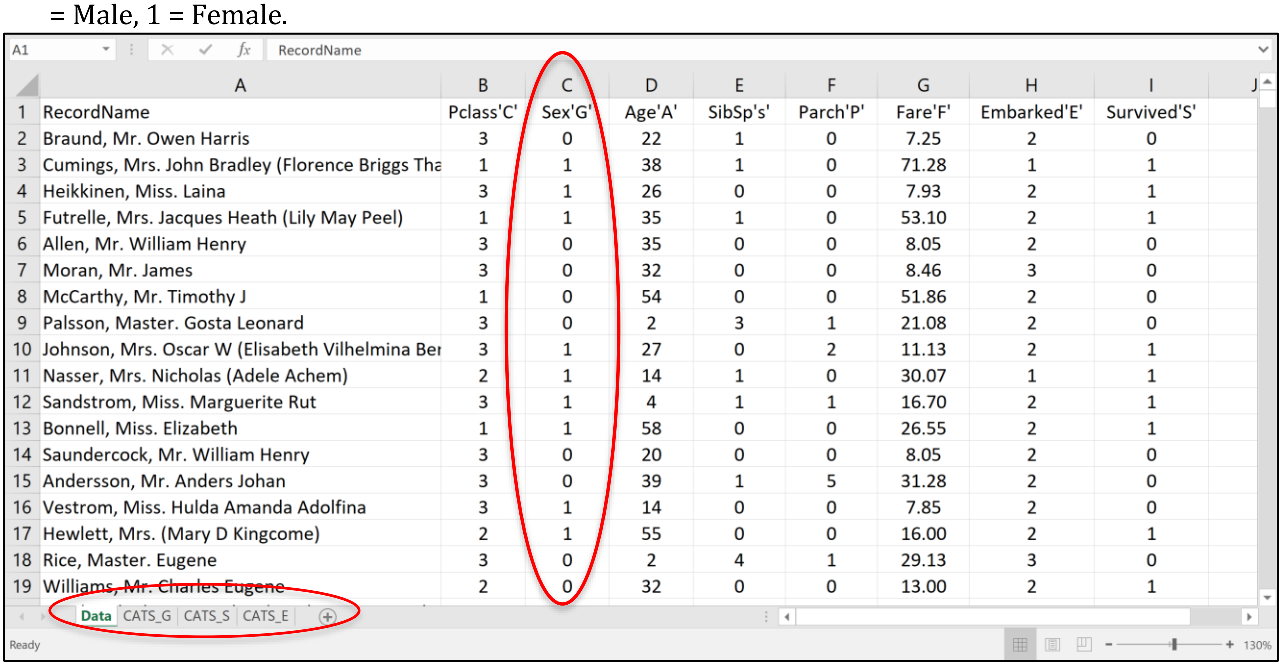
Formatting an Excel data file for upload to the Polynary Application

File must be a rectangular Excel file in the **.xlsx** format (Excel 2007 or later) with no empty cells or non-numeric values (except in the Header row or Column for *[Record Names](#)*). Note: **.xls** files must be saved as **.xlsx** files before uploading.

- File should be positioned top left, i.e. outward from cell A1.
- Columns are variables; rows are cases
- Row #1 (column headers) specifies variable names and letters that represent them in Polynary strings and Polynary graphs.
  - Column A may be used for the record names. *The column header must read*: RecordName.
  - Variable names are descriptive words or phrases followed by a unique letter symbol.
  - Variable letter symbols can be both upper and lower case and *must be enclosed in single quotes.* It is usually helpful in the interpretation of Polynary graphs if you select letters that are mnemonically linked to the variable they represent, e.g.
    e.g. GDP'G' or Inflation 'i', or Math Proficiency 'M'.
  - To reverse the directionality of an [ordinal variable](#) (e.g. where lower is better, like ERA in baseball, or Infant Mortality Rate for a country), insert a minus sign before the variable letter inside the single quotes. For example: ERA'-E'
- Only one worksheet can have data. If the Excel file has additional tabs with data those worksheets must be deleted. (Exception is for categorical variables – see Categorical Variables below).

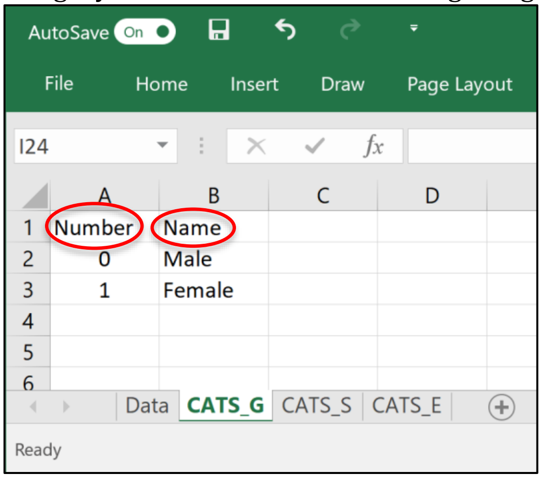
[Categorical Variables](#) may be used as independent variables, or as *[Prefilters](#)* in a *[Polynary Model](#)* to display the data by categories such as Gender or Ethnicity. *Prefilters must be identified and set up in separate tabs (worksheets) in the Excel input file before it's uploaded to Polynary.*

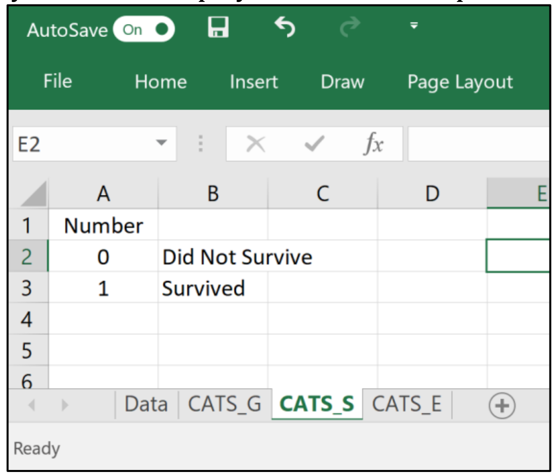
- Categories are numerically coded in the data worksheet e.g. Female = 1, Male = 0.
- Prefilters are set up in the Excel workbook by creating a separate Tab (worksheet) for each categorical variable. Each Tab is labeled CATS_*, where * represents the letter symbol representing that categorical variable.
- The example below is the Titanic Survival data set found on the Kaggle website: https://www.kaggle.com/mohitjlr/titanic/data. The categorical variable, Sex, in column C is encoded 0 = Male, 1 = Female.



	A	B	C	D	E	F	G	H	I
1	RecordName	Pclass'C'	Sex'G'	Age'A'	SibSp's'	Parch'P'	Fare'F'	Embarked'E'	Survived'S'
2	Braund, Mr. Owen Harris	3	0	22	1	0	7.25	2	0
3	Cumings, Mrs. John Bradley (Florence Briggs Tha	1	1	38	1	0	71.28	1	1
4	Heikkinen, Miss. Laina	3	1	26	0	0	7.93	2	1
5	Futrelle, Mrs. Jacques Heath (Lily May Peel)	1	1	35	1	0	53.10	2	1
6	Allen, Mr. William Henry	3	0	35	0	0	8.05	2	0
7	Moran, Mr. James	3	0	32	0	0	8.46	3	0
8	McCarthy, Mr. Timothy J	1	0	54	0	0	51.86	2	0
9	Palsson, Master. Gosta Leonard	3	0	2	3	1	21.08	2	0
10	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Ber	3	1	27	0	2	11.13	2	1
11	Nasser, Mrs. Nicholas (Adele Achem)	2	1	14	1	0	30.07	1	1
12	Sandstrom, Miss. Marguerite Rut	3	1	4	1	1	16.70	2	1
13	Bonnell, Miss. Elizabeth	1	1	58	0	0	26.55	2	1
14	Saundercock, Mr. William Henry	3	0	20	0	0	8.05	2	0
15	Andersson, Mr. Anders Johan	3	0	39	1	5	31.28	2	0
16	Vestrom, Miss. Hulda Amanda Adolfina	3	1	14	0	0	7.85	2	0
17	Hewlett, Mrs. (Mary D Kingcome)	2	1	55	0	0	16.00	2	1
18	Rice, Master. Eugene	3	0	2	4	1	29.13	3	0
19	Williams, Mr. Charles Eugene	2	0	32	0	0	13.00	2	1

- Within each tab the Headers must be: “Number” in A1 representing the numeric codes for each category; and “Name” in B1 denoting category names, as displayed in the 2 examples below.

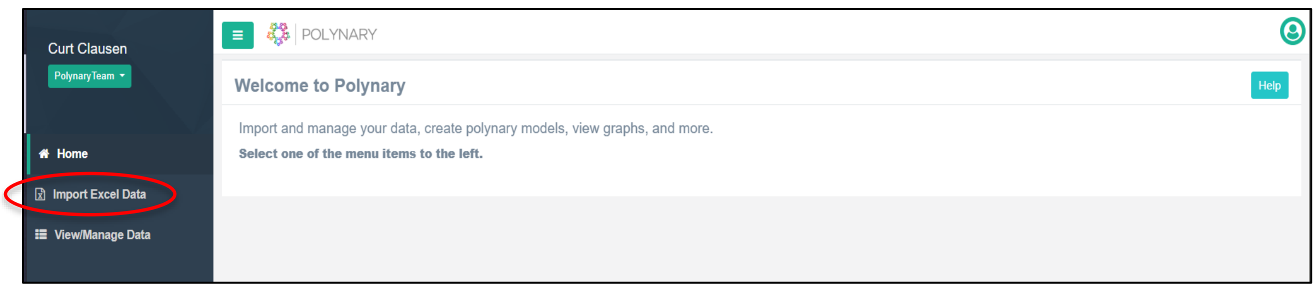




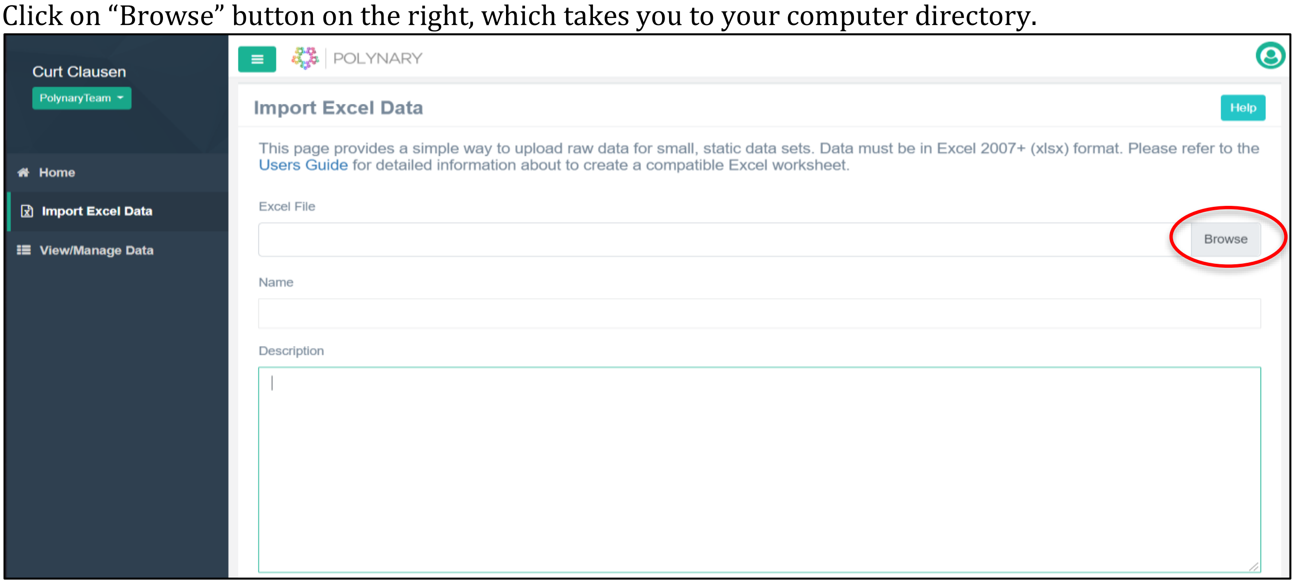
Check for errors before importing. Once the Excel workbook formatting is complete, it’s useful to check for empty or non-numeric data cells, duplicate variable letter symbols, missing or misplaced single quotation marks, or Stray input in cells outside the rectangular data.

Uploading an Excel Data Set

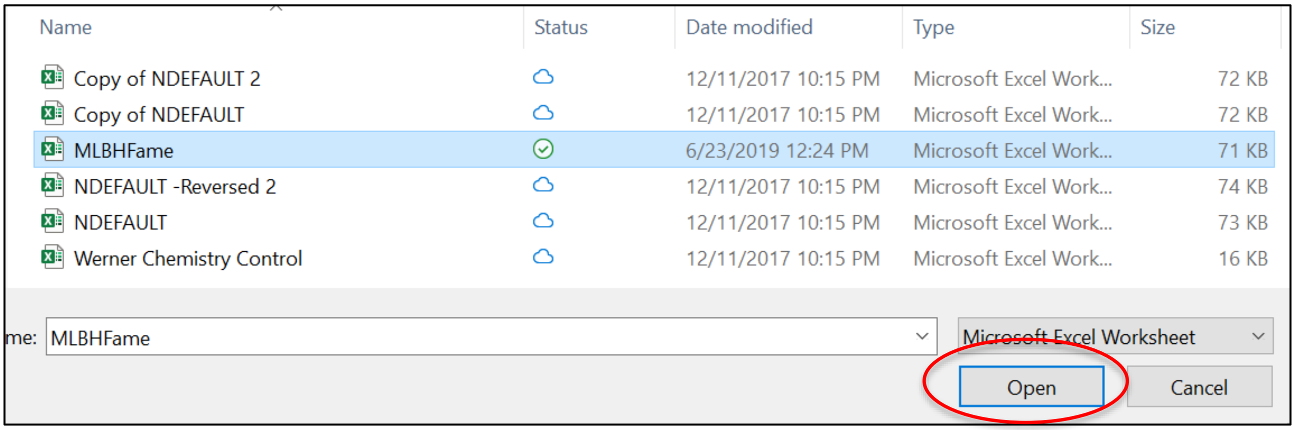
To upload an Excel Data file, sign in to your Polynary account, and from the home screen, click on “Import Excel Data” from the menu on the left.



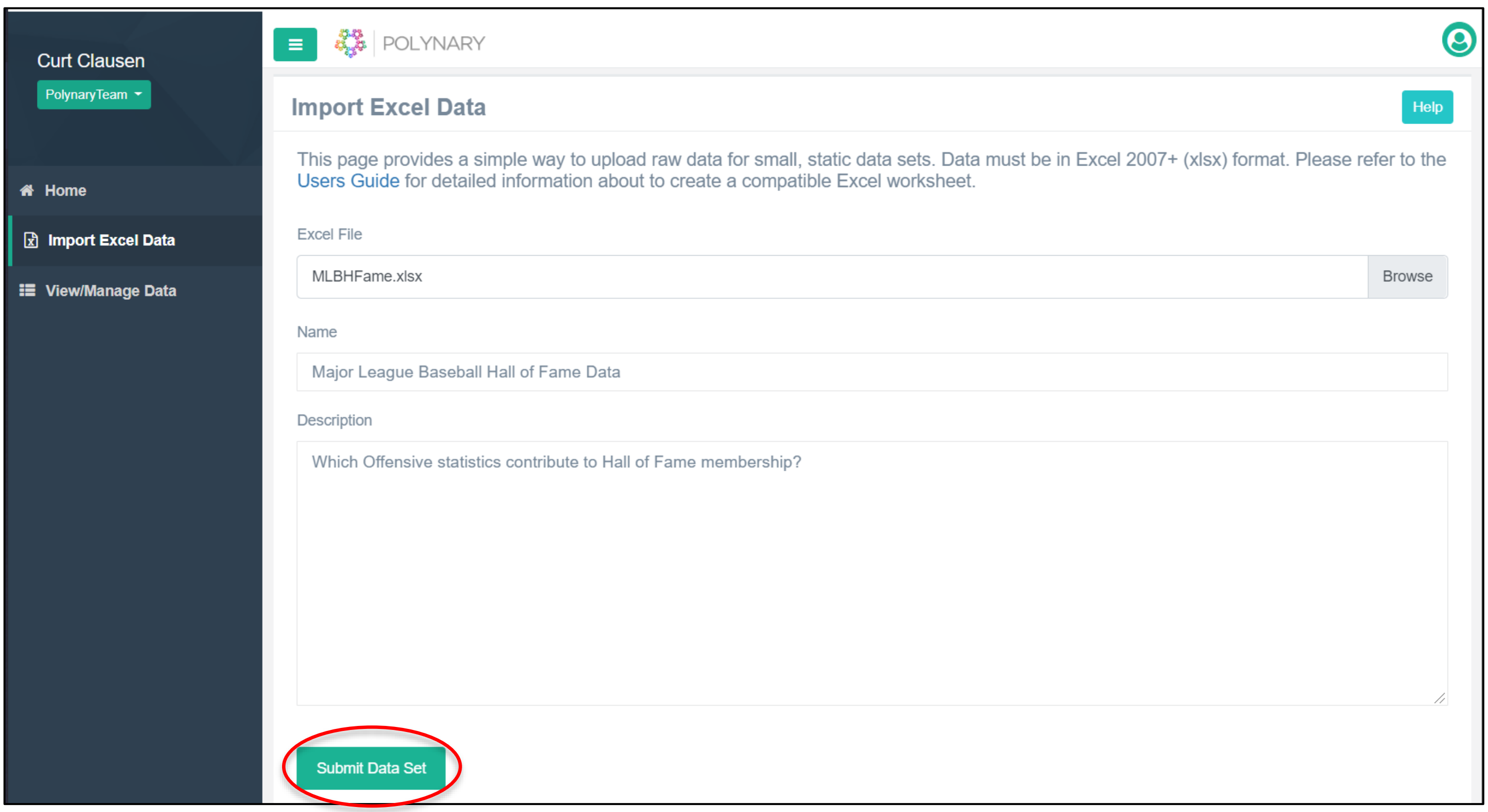
Click on “Browse” button on the right, which takes you to your computer directory.



Select the file to be uploaded and click “open”.

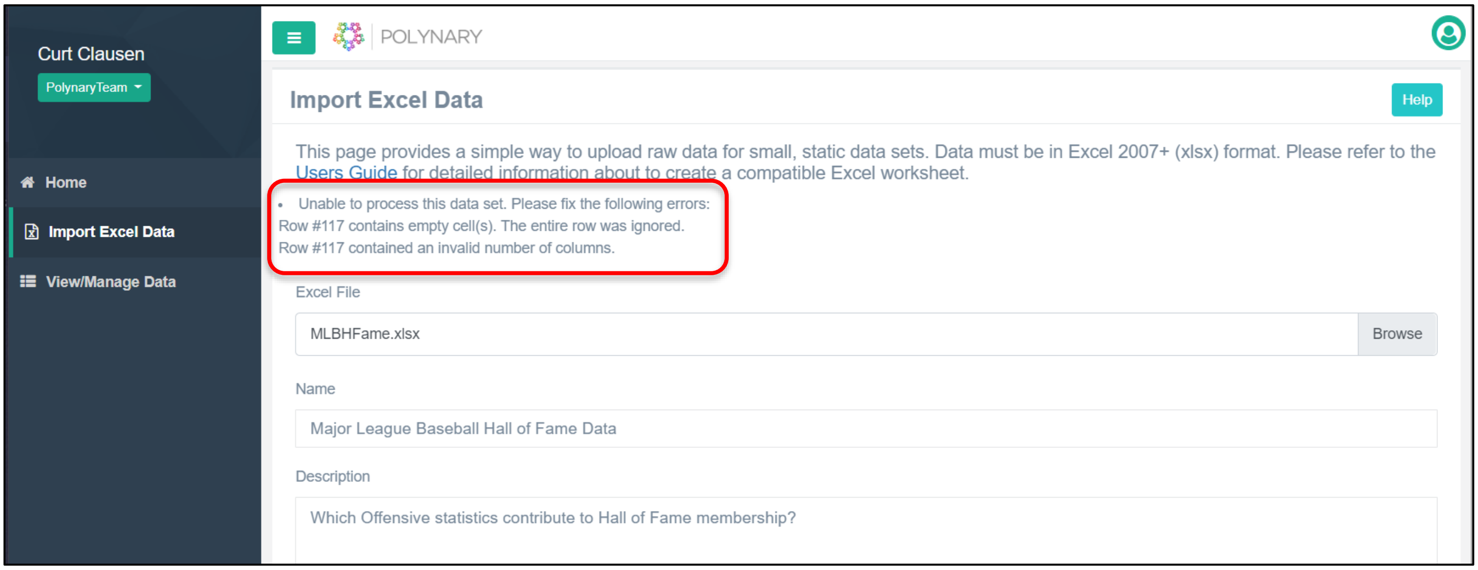


Provide a Name and Description and click on the “Submit Data Set” bar to upload the selected file.



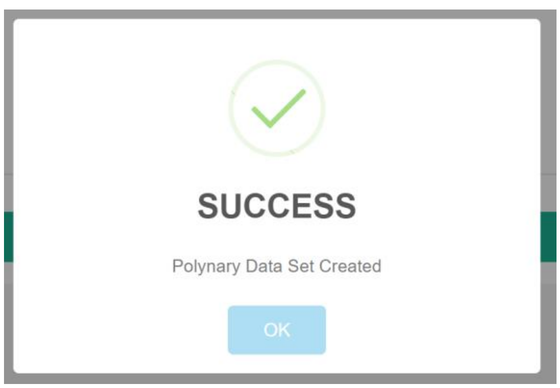
Upload will take a few seconds depending on the size of your data set.

Error Messages. If there are any problems with excel file, such as missing or non-numeric data, the application will display them as *Exceptions,* and identify the row(s) with the issues.

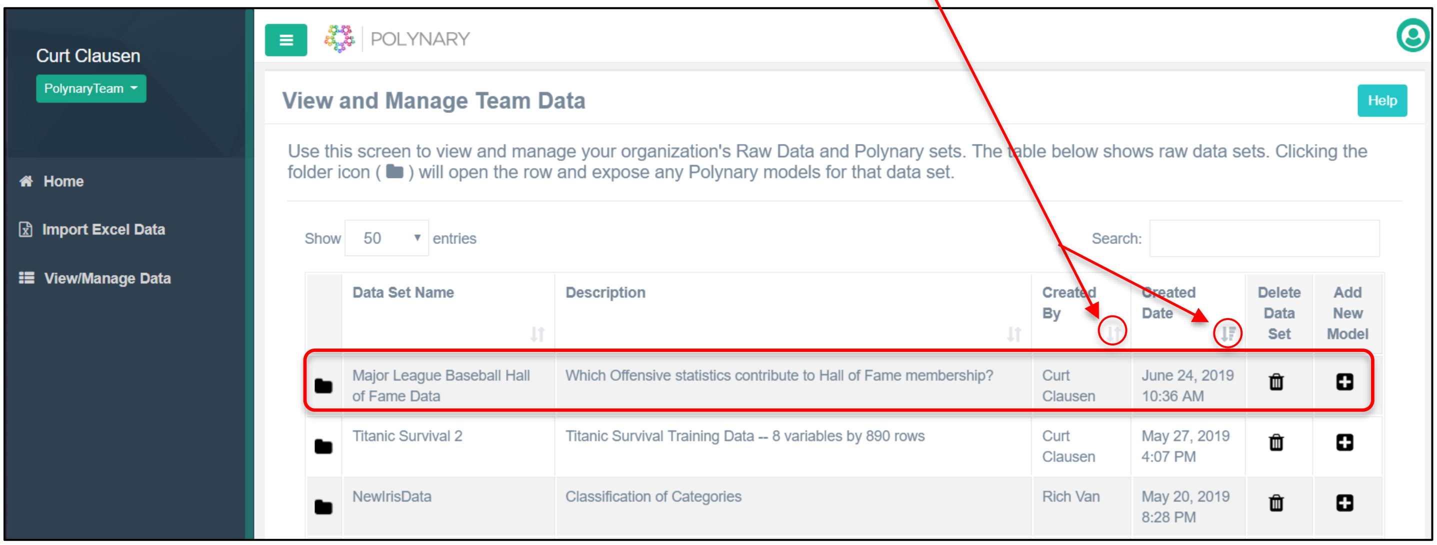


You will need to reopen the excel file, correct the identified errors, save and close the file. Then return to the Polynary application and re-upload the file.

Success! A notification appears when the file is successfully uploaded. Click "OK"

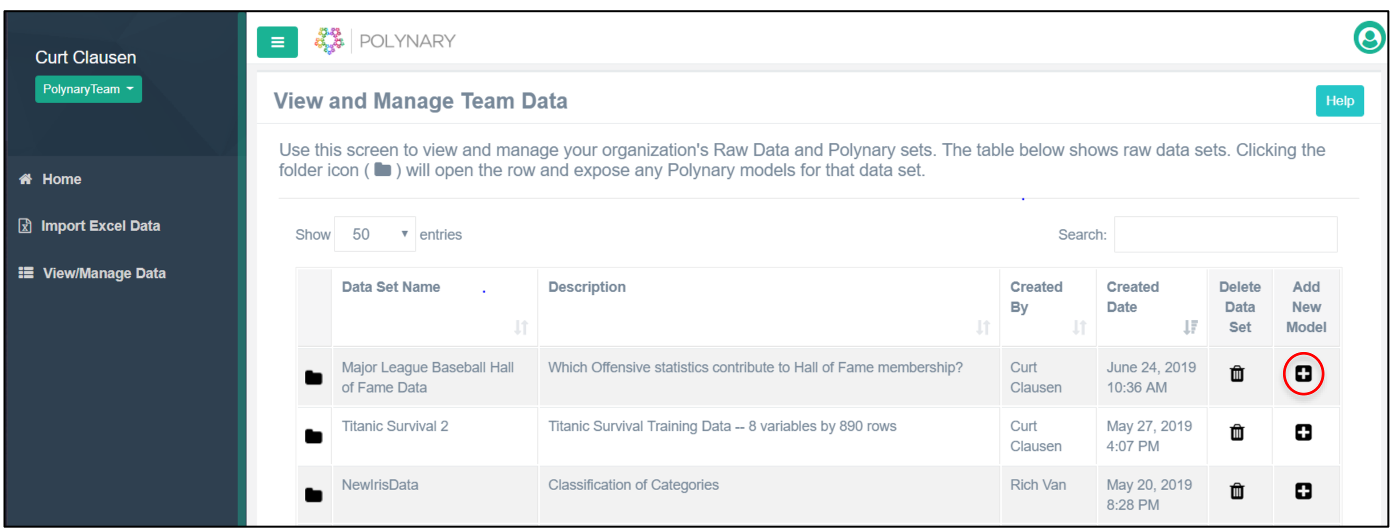


Your *[Team Data Library]* is displayed with the newly imported data set at the top. You can reorder the data sets either by "date created" or "Created by" By clicking on the arrows below those labels.



Creating a Polynary Model

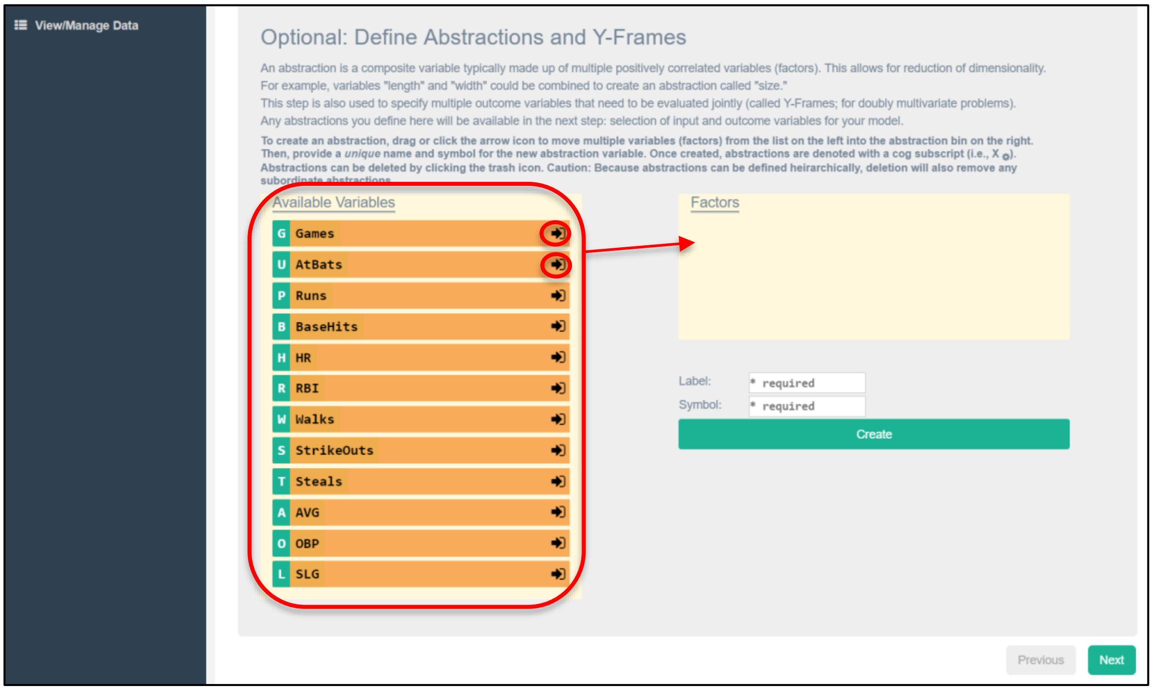
Congratulations! You are now ready to create your first Polynary visualization. To launch the Model Builder, click on the "**+**" sign under "Add New Model" at the far right of the row.



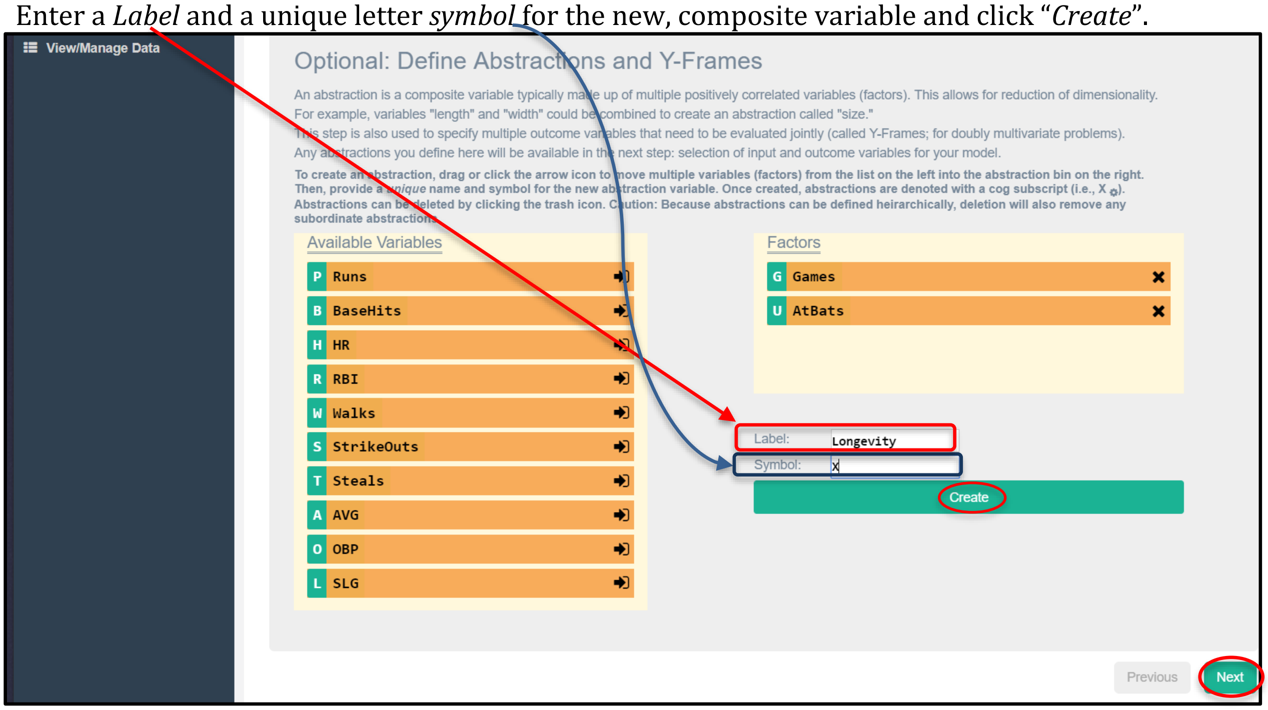
You will be guided through the Model Builder's series of screens, first, to create any *abstract* or *composite variable*, and then to select the independent, dependent and "*pre-filtering*" variables.

Creating Abstract Variables. Abstractions are multi-dimensional constructs channeled into a single continuous variable through a joint consideration of its multiple dimensional values. For example, variables "length" and "width" could be combined to create an abstraction called "size." On the first screen you have the option of creating one or more abstract variables. Once defined, they can be selected as either input or output variables in the next screen of the Model Builder.

To create an abstract variable, select the variables to be combined from Available Variables by clicking on the arrows. The selected variables will move from the "Available Variables" panel to the "Factors" canvas for this abstraction.



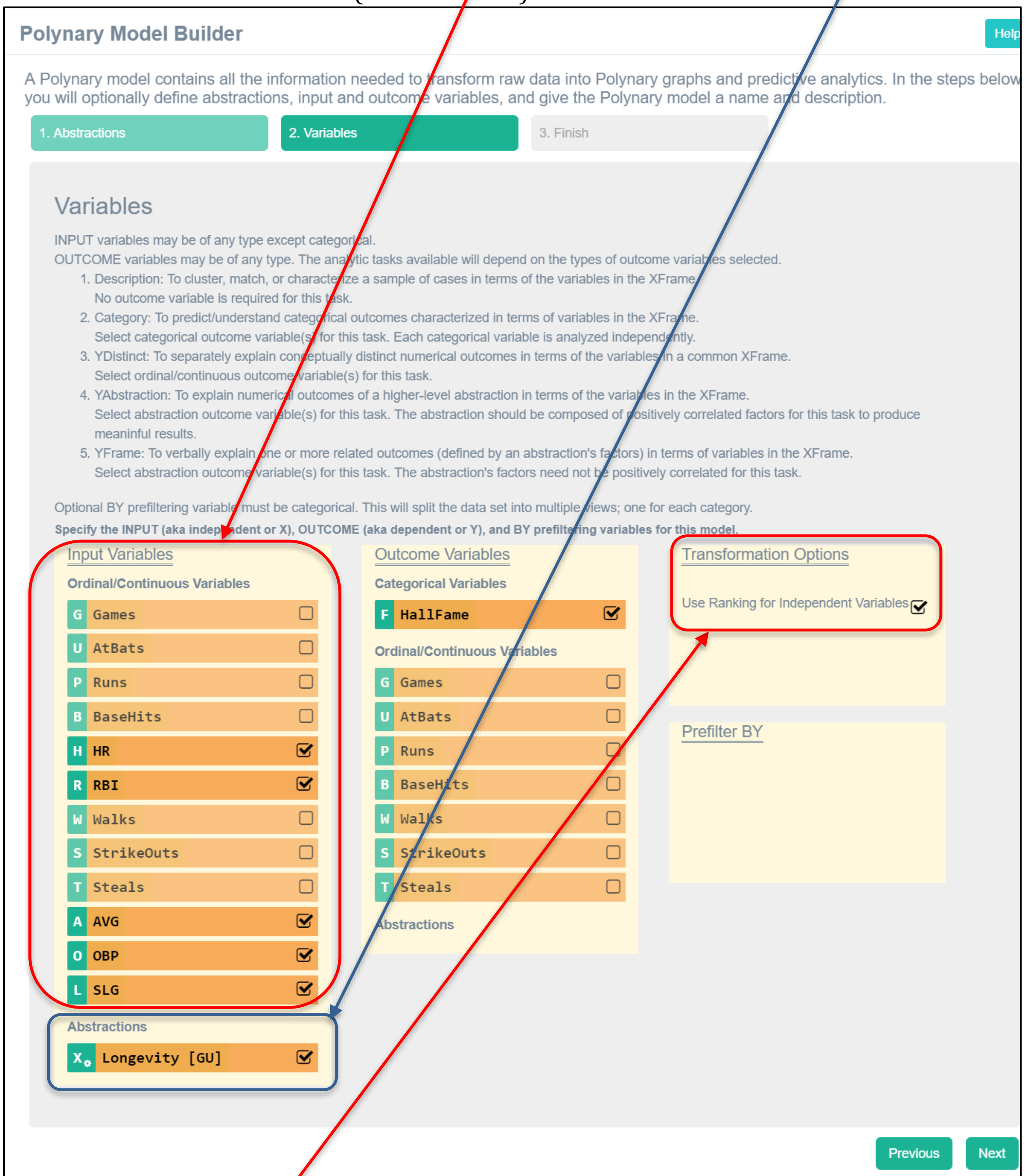
Enter a *Label* and a unique letter *symbol* for the new, composite variable and click "*Create*".



This process may be repeated to create multiple abstract or composite variables, or may be skipped if none are intended. When finished click "Next".

Selecting Independent and Dependent Variables

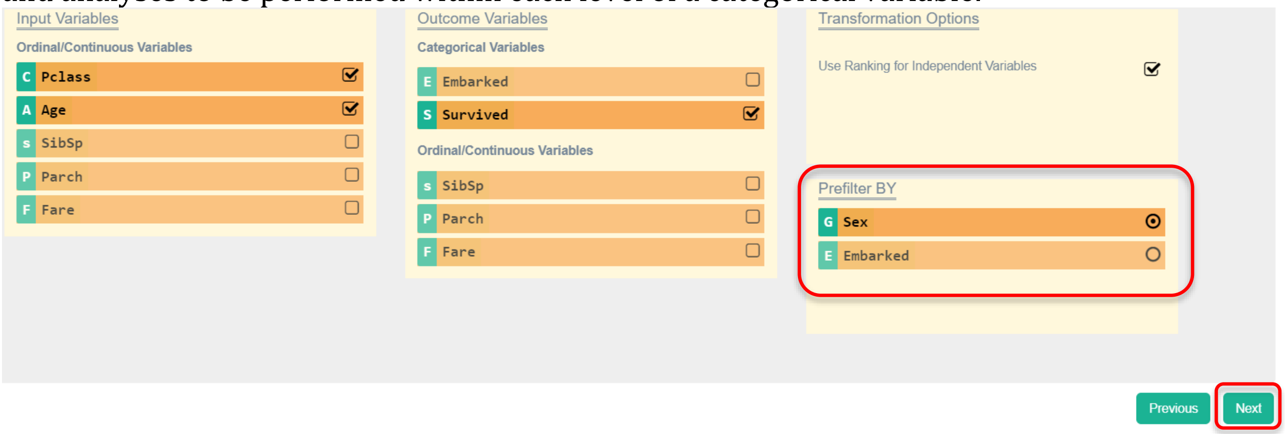
Independent variables are grouped as *Ordinal*/*Continuous Variables*, and *Abstractions* if any have been created. (Note that the composite variable "Longevity", labeled X, is included as an Abstraction Independent Variable choice). As you click in a check box to select an Input variable, it disappears from the menu of Outcome variables (and vice versa).



Transformation Options allows you to express the data values in terms of their location relative to scale endpoints (percent of scale) or relative to the other data values (percentile rank). If Ranking is selected, tied values receive the same rank. The default transforms the data as *percent of scale*. Checking the box creates a ranked transformation of the data. These scaling choices set the interpretive context for the graphs and analytic results. In general Ranking more effectively draws distinctions between objects, but it is often useful to compare both views of your data.

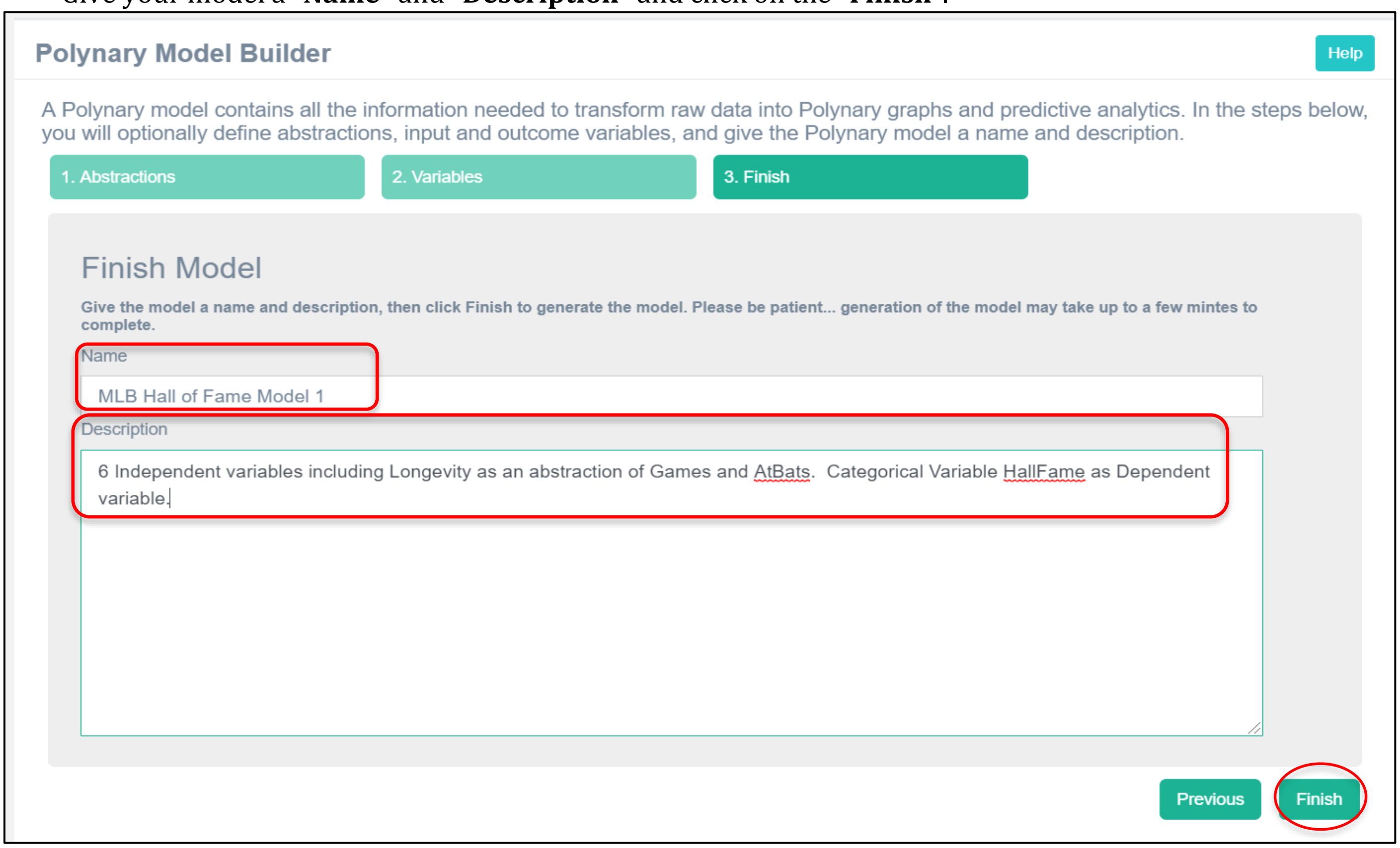
Selecting Prefilters in Polynary Model Builder

After selecting your independent and dependent variables in Model Builder, the unused *categorical variables* are displayed in the "**Prefilter BY"** canvas. Prefiltering allows the graphing and analyses to be performed within each level of a categorical variable.

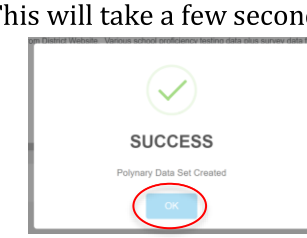


Click the radio button of any categorical variables you want filter by, and then click "**Next**".

Give your model a "**Name"** and "**Description"** and click on the "**Finish**".

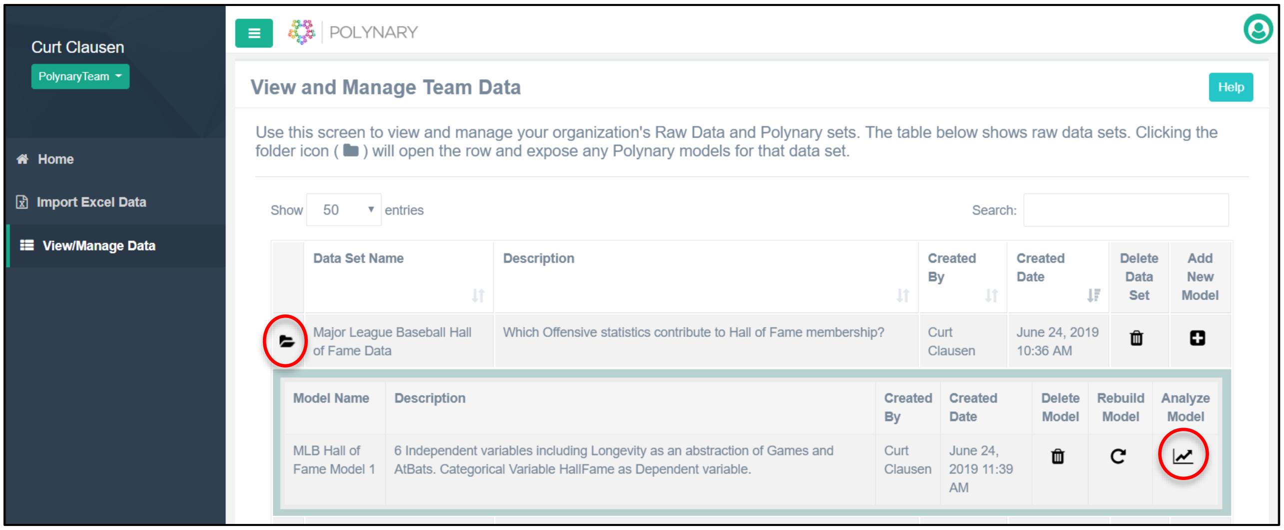


The program now transforms the numeric vectors in the Excel data into a *Polynary Strings* – letter sequences that locate each record in the Polynary fractal coordinate system. This will take a few seconds depending on the size of the data set. When done the a "***SUCCESS"*** message will be displayed.



Click "**OK**" and your Team Data Library will be displayed. Congratulations. You are ready to begin exploring your data!

Each data set will be displayed with a folder icon at the left. Individual models are displayed underneath by clicking on the folder. To view the Polynary graph of the model, click on the graph icon on the right and...



the graph will be displayed in a new browser tab:

Working with Polynary Graphs

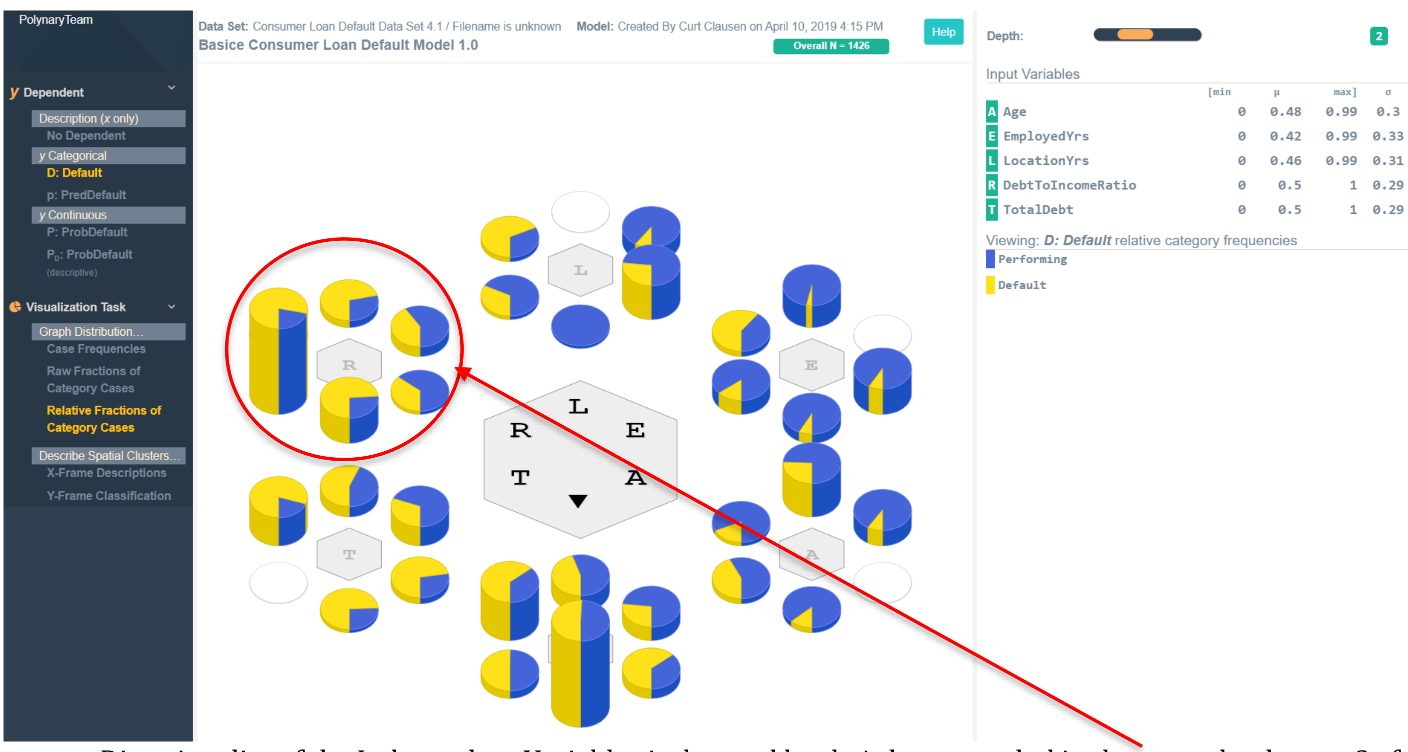
Polynary graphs consist of regular polygons with one more side than there are independent variables – that is, a data model with N independent variables will display as a polygon with N+1 sides.

- Each vertex is associated with an independent variable. The orientation of the Independent Variables is denoted by their letter symbol in the central polygon.
- Objects with extreme amounts of a single feature are located closest to the associated vertex. Objects that are more equal in terms of their input-variable features map close to the center.
- Objects with low values of all independent variables map to region "▼", which is oriented by convention at the bottom of the graph, and is analogous to the Origin of the Cartesian, X,Y coordinate system.
- The initial rendering displays the graph at "resolution 3", i.e. Polynary strings 3 letters in length.

To illustrate the features and options in Polynary we'll use a Consumer Loan Default data set as an example. The model consists of 5 input or "X" variables (*X-Frame*):

- **A** Age
- **E** EmploymentYrs (Years at current employer)
- **L** LocationYrs (Years at current residence)
- **R** DebtToIncomeRatio
- **T** Total Debt

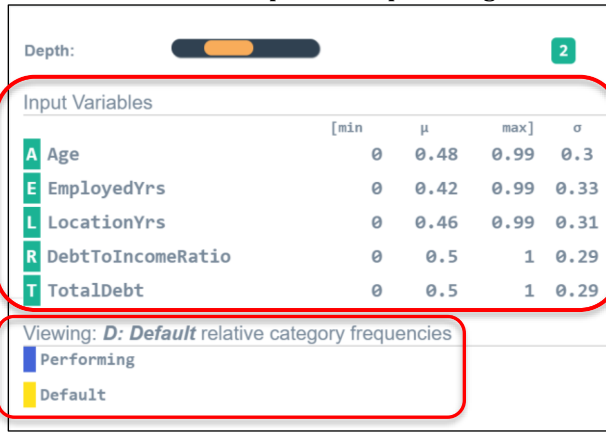
In this model there are multiple options for the Dependent "Y" Variable (*Y-Frame*). Here is a view of the binary categorical Y variable: whether the loan is *Performing* or in *Default.*



Directionality of the Independent Variables is denoted by their letter symbol in the central polygon. So for example, borrowers relatively highest in their DebtToIncome ratio are found in the "R" region on the graph.

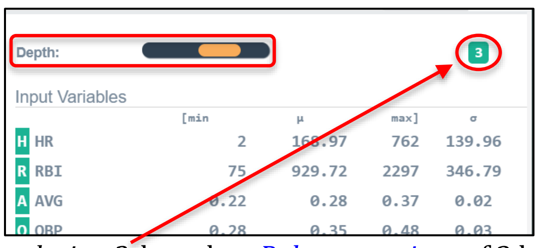
- The Legend identifies the variable names associated with each letter symbol.
- Stack pie charts denote the relative likelihood of default in each Polynary basin
- Stack height is proportional to the number of cases in that basin.

The Legend in the upper right-hand corner displays the Input Variables (X-Frame), including the letter symbol, along with its population Min, Max, Mean and Standard Deviation for each input variable. In this example the data had been re-scaled 0 – 1.0 prior to uploading to the Polynary app.



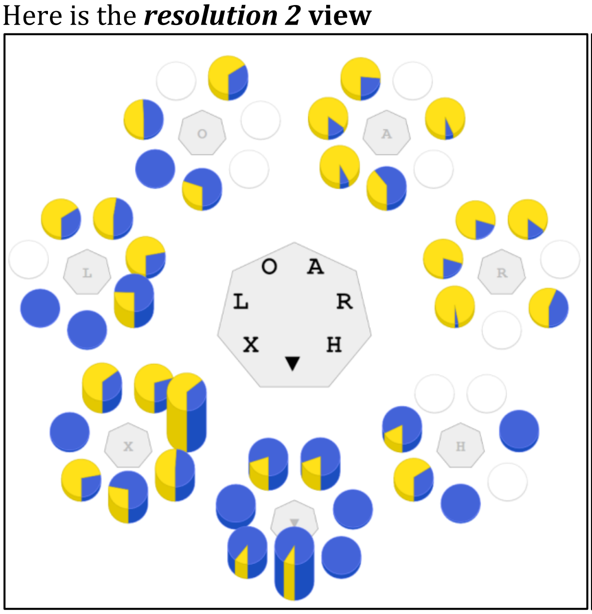
The color coded dependent (Y) variable in this view is the binary categorical variable: whether the loan is *Performing* or in *Default*.

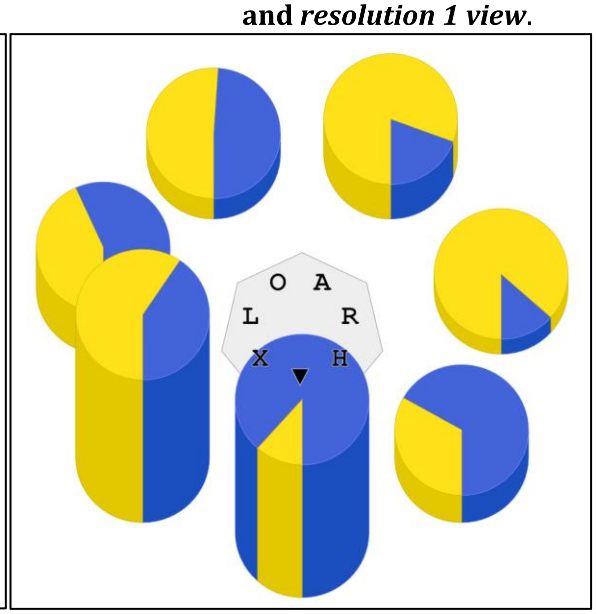
Graph Resolution A *Depth slider* changes the granularity (called *resolution* in Polynary-speak) displayed in the graph.



The default setting is *resolution 3*, based on [*Polynary strings*]() of 3 letters, and is displayed to the right of the slider. Moving the slider to the left reduces the resolution displaying a less granular view of the data. It's often helpful to view the graph at different resolutions to find the one that provides the best insight

Here is the ***resolution 2* view** **and *resolution 1 view***.





Choice of graph resolution is subjective and will depend on the data set, the questions being asked of it, and the audience for whom the analysis is created.

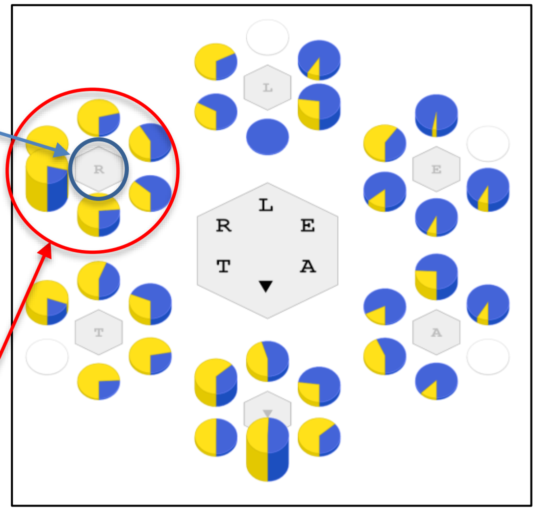
Zooming

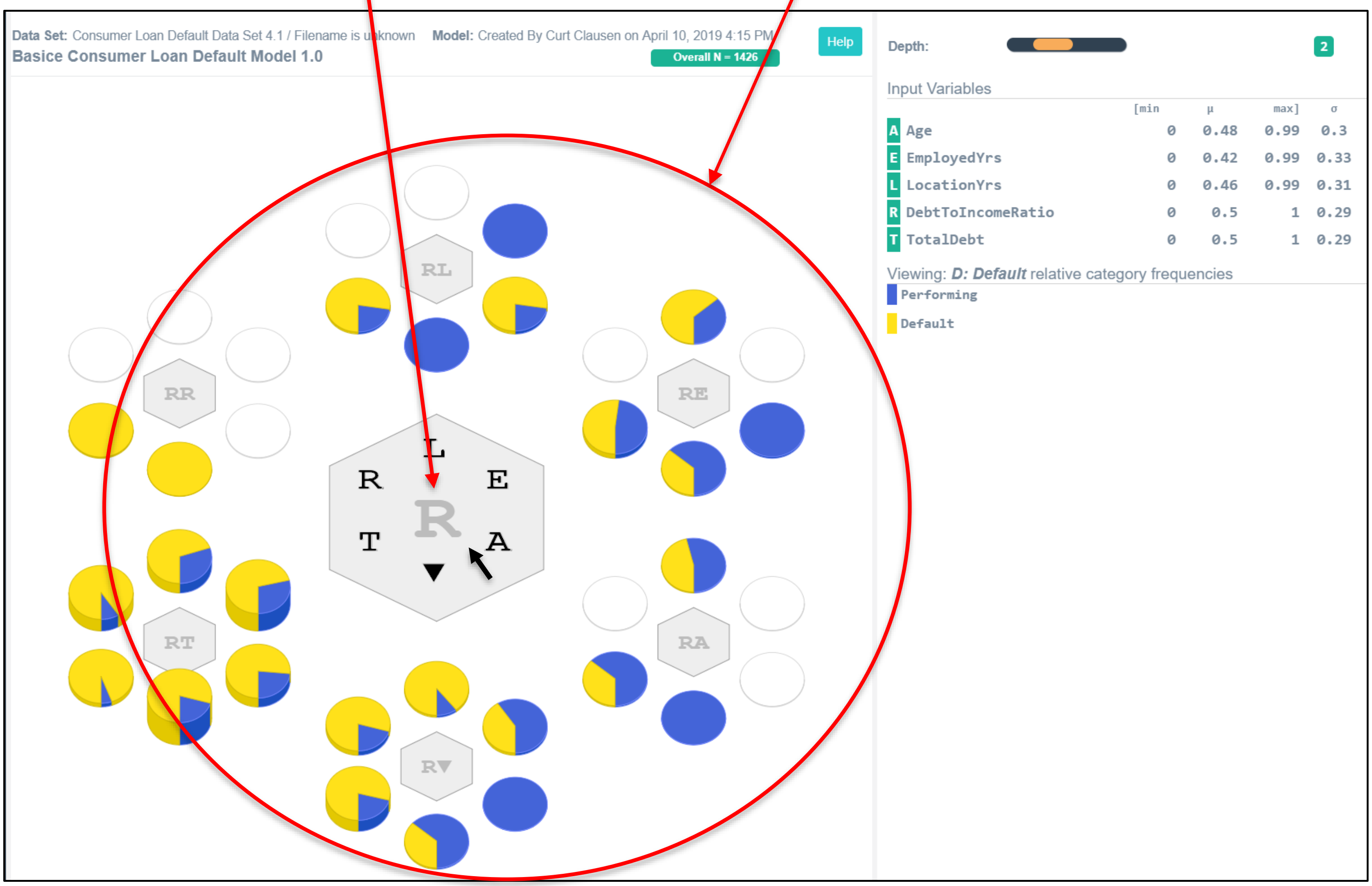
To zoom into a specific region of the graph in greater detail, click on the gray polygon *in the center of that region*.

Region "R" (DebtToIncome Ratio) is hi-lighted in the "root level" graph on the right, and rendered as "zoomed" in on in the graph below. The letter "R" is now displayed in the center of the graph indcating that we are viewing sub-region "R" only.

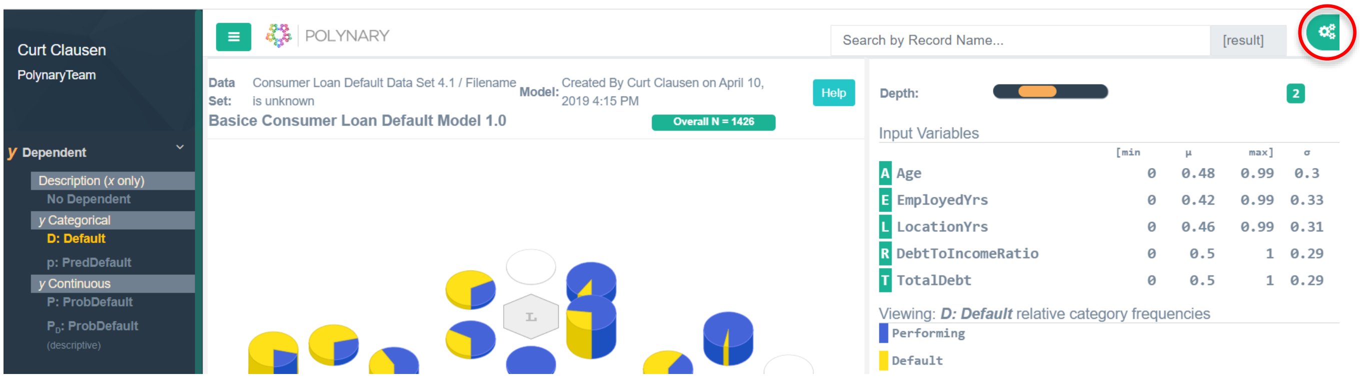
In the current version of the software 4 levels of zooming in are supported.

To reverse the process and zoom back to a higher level view, click on *the gray central polygon*.

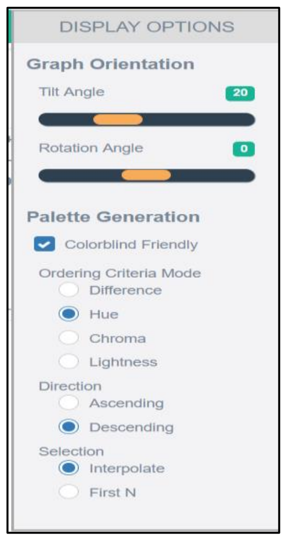




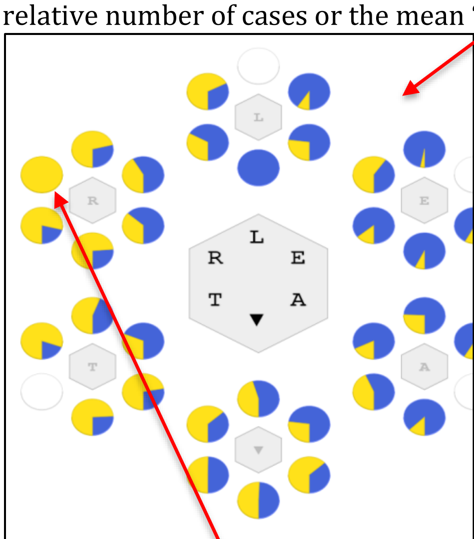
Display Options. Click on the *gear icon* to access the *tool tab,* a menu of additional display options. To hide the display options panel, click the gear icon again.

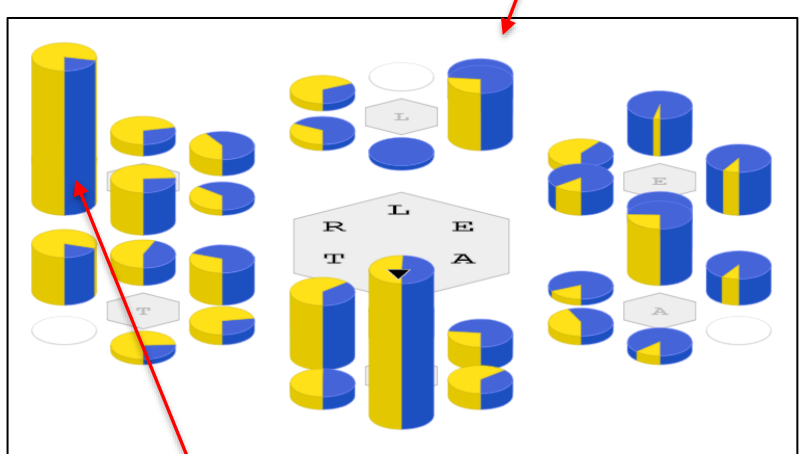


This menu pops out and provides additional tools to modify and enhance the graph's readability.



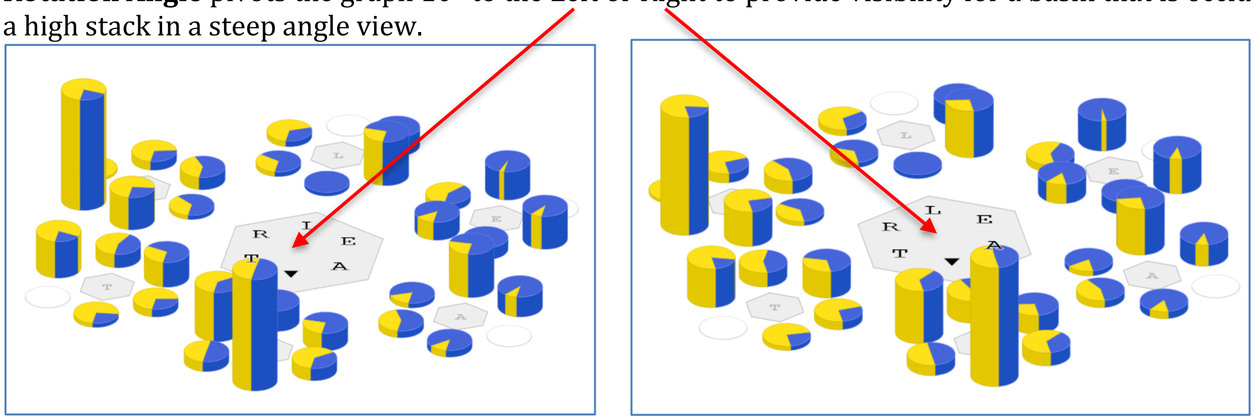
Tilt Angle changes the viewing angle from 0° for a top-down look, to as much as 60° to emphasize the relative number of cases or the mean "Y" variable value in each basin. The default setting is 20°



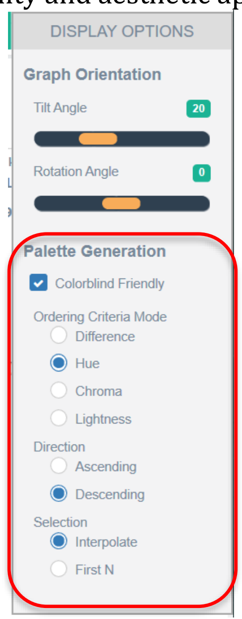


Note that basin "RR" is occluded from view by basin "RT" in the 60° view. The *Rotation Angle* slider helps solve this problem.

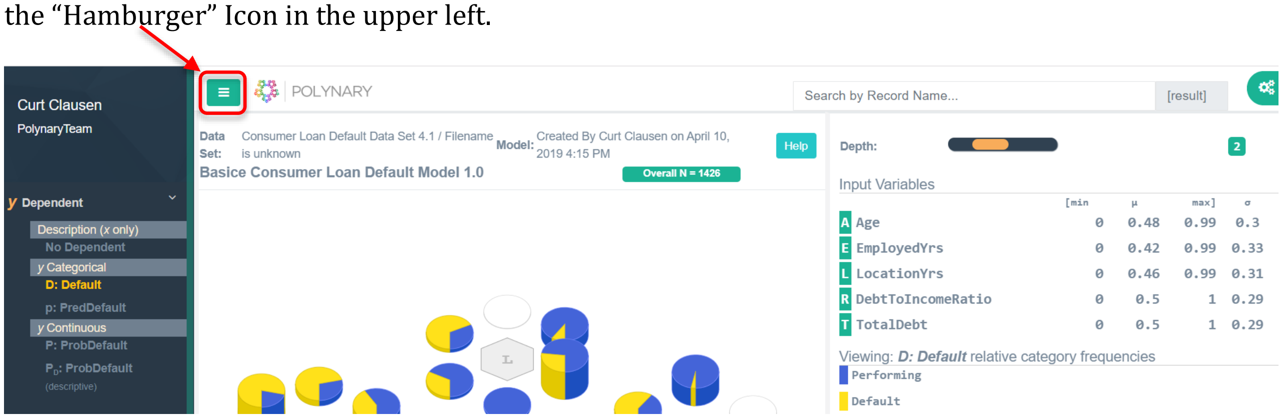
Rotation Angle pivots the graph 10° to the Left or Right to provide visibility for a basin that is occluded by a high stack in a steep angle view.



Color Palette The *Palette Generation* menu provides several color palette options. Try experimenting to find the best combination of readability and aesthetic appeal.



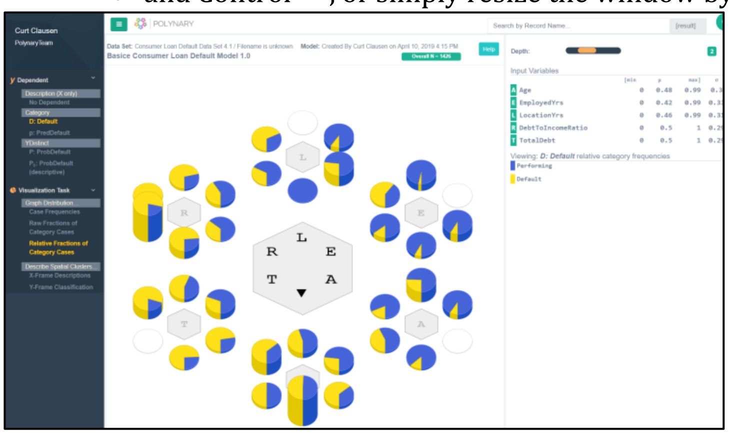
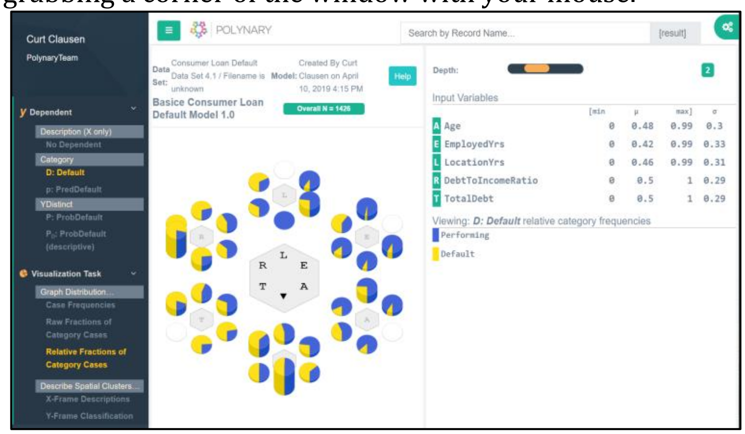
Changing the Size of the Graph To create more screen real estate for the Polynary Graph you can click on the "Hamburger" Icon in the upper left.



The left hand black vertical menu pane disappears, and the graph expands.

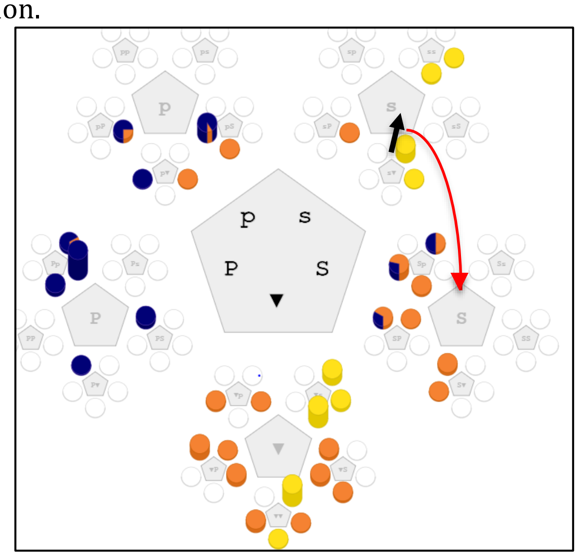
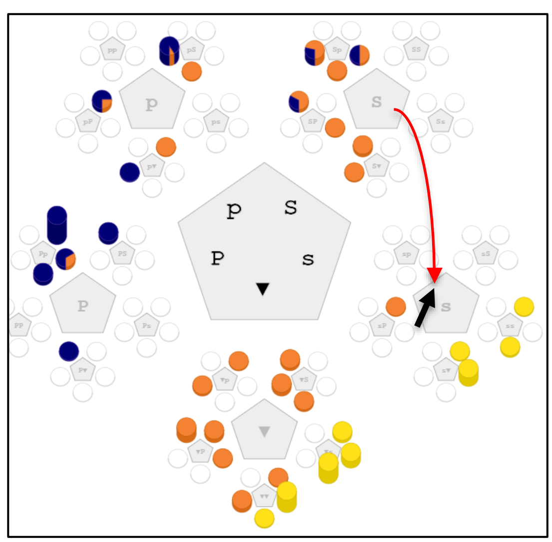
To return the left menu back into view click on the Hamburger icon again.

To *expand* or *shrink* the size of the graph relative to the legend and menu, in Chrome or Edge browser use Control "+" and Control "-", or simply resize the window by grabbing a corner of the window with your mouse.

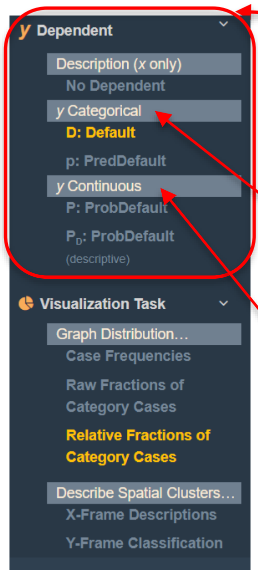



Re-ordering the Input Variables

The visual insight can sometimes be enhanced by changing the order of the input variables in the graph. To re-order the layout, simply grab the gray central polygon of the variable you want to move and drag it to its new location.

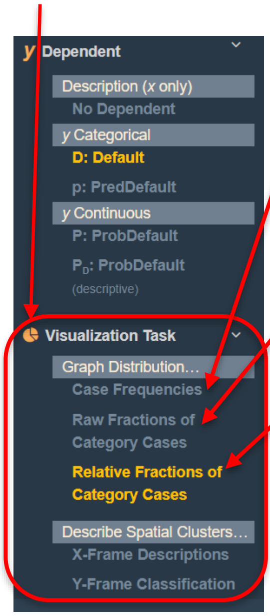
Model Selection Menu. The vertical left-hand menu allows you to select your dependent variables and analytic or visualization tasks. The current view is always hi-lighted in **gold font**.



The **y Dependent** section lists the choices for the *Y-frame* -- that is the set of outcome (dependent) variables that you created specified as Outcome Variables in the Poynary Model Builder:

- **Description (X only)** view displays the *X-frame* distribution with no outcome Y variable. The stack heights indicate the relative frequencies of the cases in each Polynary basin.
- ***yCategorical*** lists *categorical* dependent variables from those created in the Model Builder. In the example, the dependent variables are **D: Default**, and **P: PredDefault**
- **yContinuous** lists all *ordinal* or *continuous Y (output) variables* created in Model Builder. In this example:
  - **P: ProbDefault** displays the continuous Y variable *Probability of Default.* Note: this variable was part of the raw data set -- not calculated in the Polynary transform.
  - P_α:: ProbDefault (descriptive) The program automatically divides the scale into intervals based on natural data clusters within this Y-variable. This results in a set of Y-bins that provide the most concise summary of the Y-data. The graph shows the frequency of these descriptions in the data. This same set of Y-bins are used when you select the Y-frame explanation option (last option in the list on the last panel). In this context, these Y-bins represent what is being explained in terms of the X-frame. They represent 'what' is being explained while the X-frame represents 'what' is doing the explaining.

Visualization Task options are displayed in the bottom half of the menu.



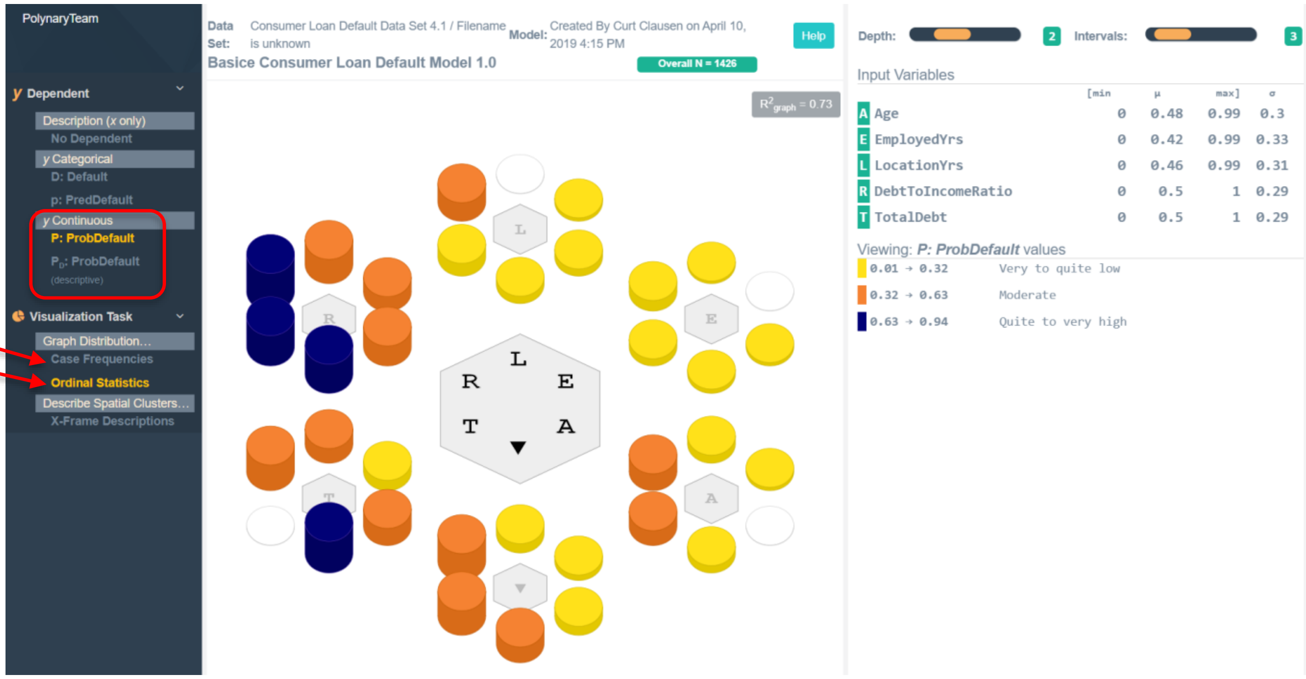
The options provided under **Graph Distribution** will depend on the type of Dependent Variable selected. If a categorical Y variable is selected the available options are:

- **Case Frequencies** displays only the X Frame as with the **Description (x only)** option under ***y* Dependent** above.
- With a *Categorical Y variable* selected, you can select either **Relative Fractions,** or **Raw Fractions of Category Cases.**
  - **Raw Fractions of Category Cases** displays the empirical distribution of the sample data and answers the question: "Of all the objects falling into a given basin, what percent belong to each Category?"
  - **Relative Fractions** is the default view and answers the question: "Given that an object falls into a certain basin, what is the likelihood that it is a member of Category X?"
- An important caveat here: Raw Fractions of Category Cases can give a highly distorted view of the data if the number of cases in each category becomes unequal. Relative Fractions of Category Cases are used to draw inferences in classification problems.

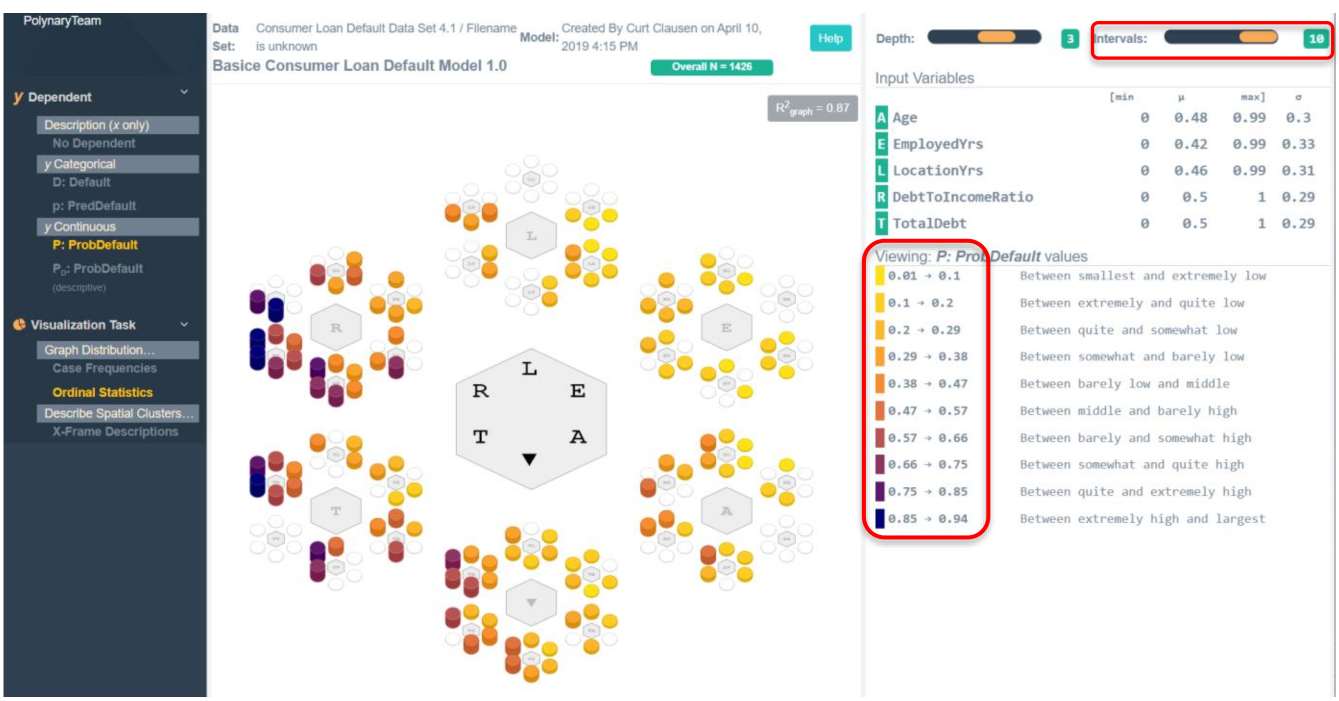
Dependent Variable Options

If a continuous Outcome variable is selected (under **YContinuous)** the **Graph Distribution** options are:

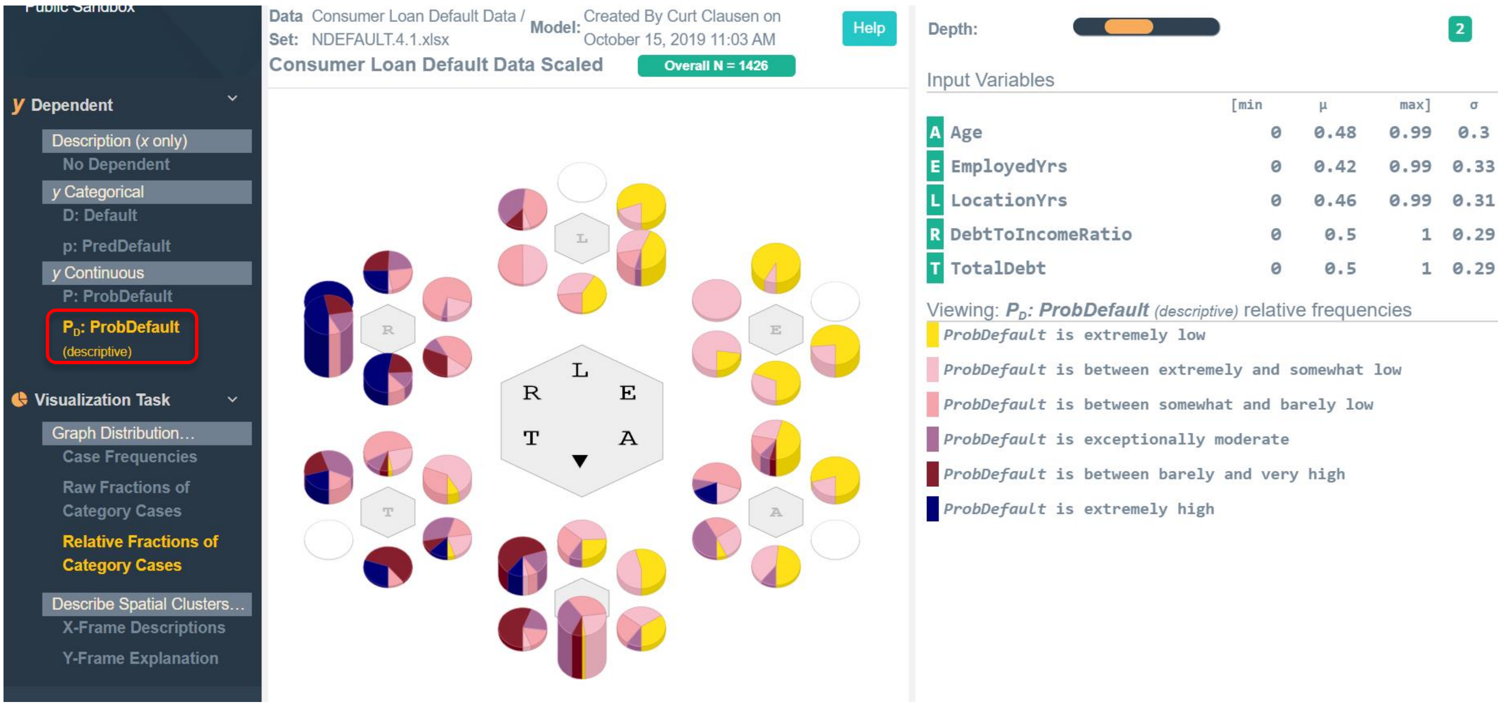
- **Case Frequencies -** display the X-frame data ignoring the outcome variable, or
- **Ordinal Statistics -** display the mean outcome of the selected Y for each X-frame basin.



If **Ordinal Statistics** is selected a new ***Intervals*** slider control appears that allows you to vary the number of intervals with which to display the dependent variable. The view below adjust both the Depth (to resolution 3) and the number of *Y variable intervals* to the maximum of 10.



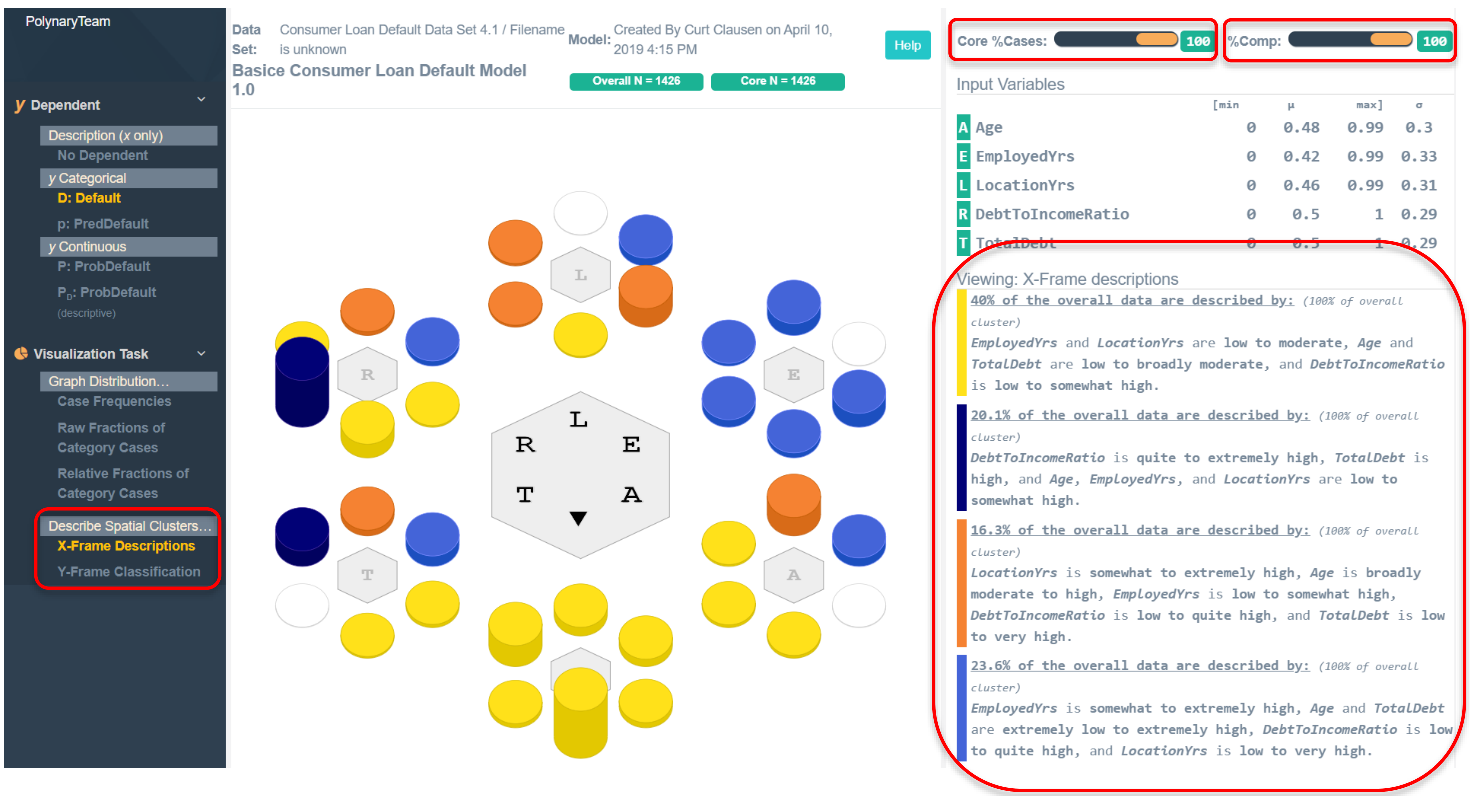
P_{α}:: ProbDefault (descriptive) The program automatically divides the continuous Y-variable into scale intervals based on natural data clusters. This results in a set of Y-bins that provide the most concise summary of the Y-data. The graph shows the frequency of these descriptions in the data. This same set of Y-bins are used when you select the Y-frame explanation option (last option in the list on the last panel). In this context, these Y-bins represent what is being explained in terms of the X-frame.



Visual Cluster Analysis and Language Description

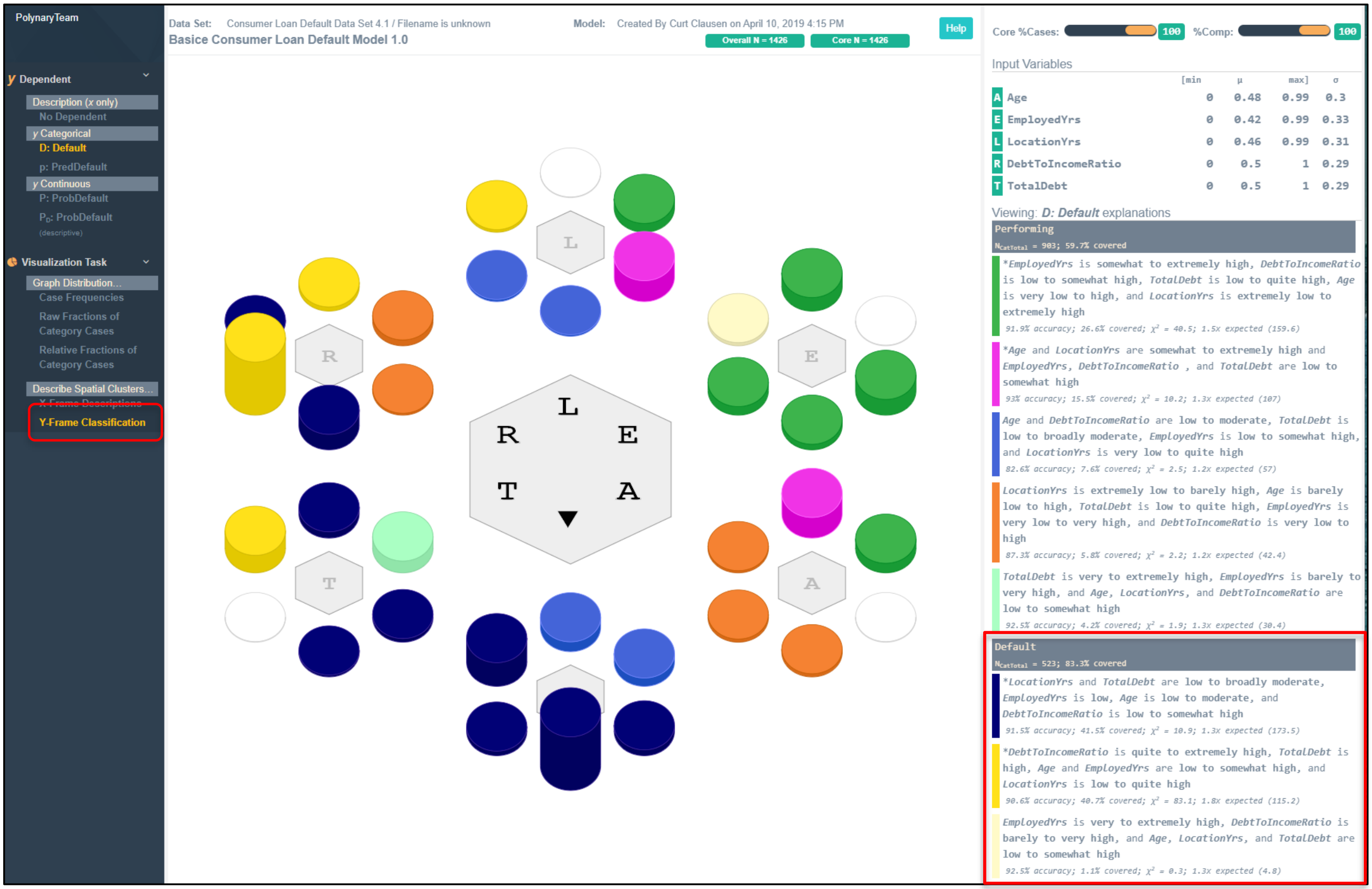
Describe [Spatial Clusters] menu offers options to display the data summarized into sets of color-coded clusters. Each cluster is a region defined as a set of spatially connected basins. *Natural language descriptions* corresponding to each cluster are shown in the legend.

- **X-Frame Descriptions** simply displays the empirical distribution of clusters in the data. Stack height represents the relative frequencies of cases in each basin.
- Sliders allow you to adjust for the amount of information displayed.
  - ***Core %cases*** controls the total fraction of the data cases described. For instance, to generate descriptions that cover the majority of cases, set the slider to 51%.
  - ***%Comp*** (Completeness) controls the percentage of the total amount of descriptive information to display in the output. Shortened descriptions that are adequate to draw the necessary distinctions are easier to remember and communicate.



Y-Frame Classification

This graph view again shows the data summarized into sets of color-coded clusters, each cluster defining a set of spatially connected basins. The corresponding cluster descriptions are shown in the legend. In this view the clusters are displayed separately in their binary Y-variable categories: *Performing,* and *Default.* In the example below, the descriptions of the dark blue, bright-yellow and pale-yellow clusters distinguish borrowers likely to default on their loans.



There are two numbers in the header of each category that provide additional details. *The first* is the number of cases in that category and *the second* is the percentage of that category's cases that are covered by the cluster descriptions beneath it.

Default

$N_{CatTotal}$ = 523; 83.3% covered

LocationYrs* and *TotalDebt* are **low to broadly moderate, *EmployedYrs* is **low**, *Age* is **low to moderate**, and *DebtToIncomeRatio* is **low to somewhat high**

91.5% accuracy; 41.5% covered; χ^2 = 10.9; 1.3x expected (173.5)

DebtToIncomeRatio* is **quite to extremely high, *TotalDebt* is **high**, *Age* and *EmployedYrs* are **low to somewhat high**, and *LocationYrs* is **low to quite high**

90.6% accuracy; 40.7% covered; χ^2 = 83.1; 1.8x expected (115.2)

EmployedYrs is **very to extremely high**, *DebtToIncomeRatio* is **barely to very high**, and *Age*, *LocationYrs*, and *TotalDebt* are **low to somewhat high**

92.5% accuracy; 1.1% covered; χ^2 = 0.3; 1.3x expected (4.8)

There are four additional statistics below each cluster description. *Accuracy* reflects how well a description connotes the distribution of data falling within a denoted region of space . The adverb phrases within a description are each selected to maximize accuracy—that is, given how people understand adverbs, no better choice can be identified. Of the 523 defaulters in the sample, 83.3% are covered by three descriptions. the first description contains 41.5% of defaulters, 40.7% are covered in the second, and the remaining 1.1% in the third description. The chi-square values are for individual cells in a two-way table: the columns are Performing and Default and the rows are the complete list of verbal descriptions for both categories. Each chi-square value above 3.86, with one degree of freedom; is significant at 0.05--inferring that a disproportionately high number of cases are found in this cell relative to random chance. Another way to report this out is how many times larger the observed number of cases in this cell is, relative to the expected number, if the rows and columns are independent of one another. For example, in the dark blue cluster we expected 173.5 cases, but observed 1.3 times that number.

Default

$N_{CatTotal}$ = 523; 83.3% covered

LocationYrs* and *TotalDebt* are **low to broadly moderate, *EmployedYrs* is **low**, *Age* is **low to moderate**, and *DebtToIncomeRatio* is **low to somewhat high**

91.5% accuracy; 41.5% covered; χ^2 = 10.9; 1.3x expected (173.5)

DebtToIncomeRatio* is **quite to extremely high, *TotalDebt* is **high**, *Age* and *EmployedYrs* are **low to somewhat high**, and *LocationYrs* is **low to quite high**

90.6% accuracy; 40.7% covered; χ^2 = 83.1; 1.8x expected (115.2)

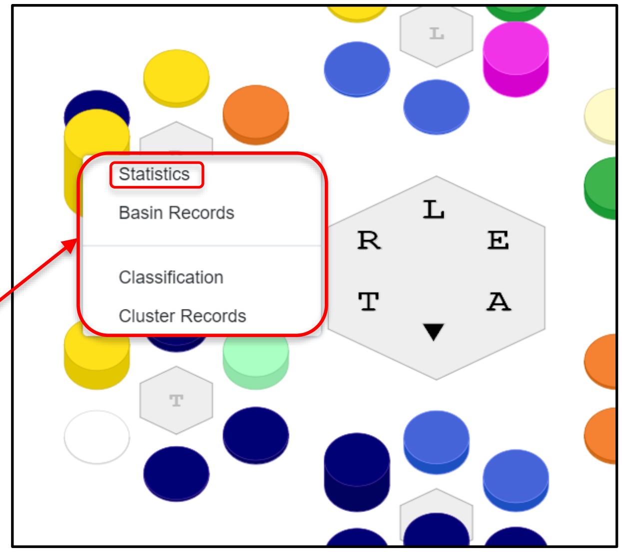
EmployedYrs is **very to extremely high**, *DebtToIncomeRatio* is **barely to very high**, and *Age*, *LocationYrs*, and *TotalDebt* are **low to somewhat high**

92.5% accuracy; 1.1% covered; χ^2 = 0.3; 1.3x expected (4.8)

Viewing Basin Level Summary Statistics, Descriptions and Records.

The Polynary String and the number of cases in a Polynary basin or a region are displayed in a pop-up by hovering the cursor over the gray polygon in the center of a region.

To view basic level statistics, right click on a basin and a pop-up menu appears:



Clicking on the "*Statistics'* option produces a table with the basin Label (RT in this example), the number of cases (N), the Min, Max, & Mean (μ) values and Standard Deviation (σ) of the Independent Variables and (in this example) the *Continuous* Dependent Variable in that basin.)

Basin 'RT' STATISTICS N = 215

Independent				
Variable	**Min**	**μ**	**Max**	**σ**
Age	0	0.37	0.88	0.2
EmployedYrs	0	0.34	0.94	0.26
LocationYrs	0	0.34	0.85	0.22
DebtToIncomeRatio	0.58	0.87	1	0.1
TotalDebt	0.52	0.74	0.98	0.12

Dependent: ProbDefault (P)				
Variable	**Min**	**μ**	**Max**	**σ**
ProbDefault	0.28	0.67	0.94	0.17

For *Categorical* Dependent Variables, the Count, Raw and Adjusted **%** Distribution of the given basin are shown. The %Adjusted accounts for the fact that there are more cases with Performing Loans than Default Loan cases in the sample. This example shows that people in basin RT are much more likely to Default

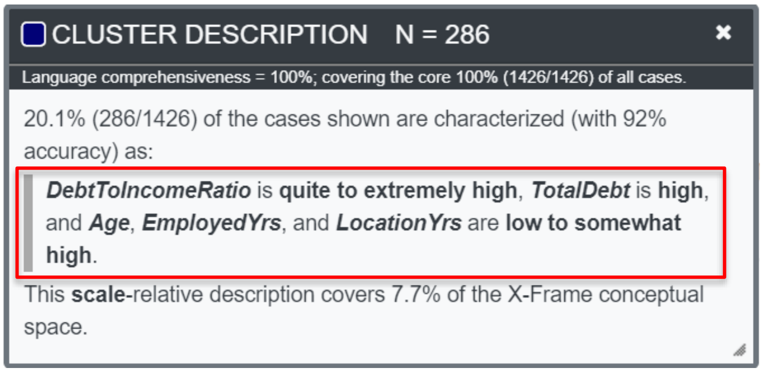
Basin 'RT' STATISTICS N = 215

Independent				
Variable	**Min**	**μ**	**Max**	**σ**
Age	0	0.37	0.88	0.2
EmployedYrs	0	0.34	0.94	0.26
LocationYrs	0	0.34	0.85	0.22
DebtToIncomeRatio	0.58	0.87	1	0.1
TotalDebt	0.52	0.74	0.98	0.12

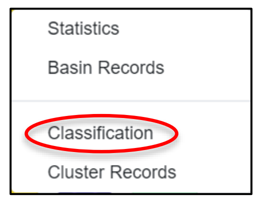
Dependent: Default (D)			
Category	**Count**	**%**	**%Adjust**
Performing	68	31.63	21.13
Default	147	68.37	78.87

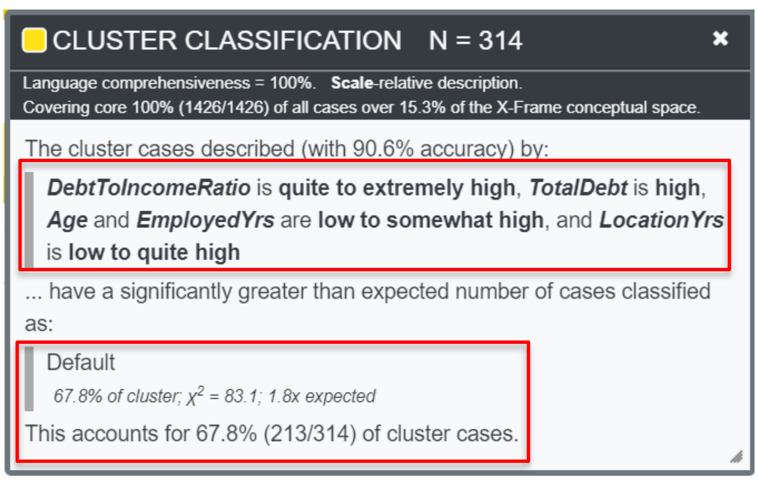
To display the summary statistics for the entire population **right click** on the large gray polygon in the center of the graph.   To display the summary statistics for a given sub-region, right click on the gray central polygon of that region.

In Cluster mode, **X-Frame Descriptions**, the menu includes a *Description* option which displays a *verbal description* of the total cluster – that is of all basins of that same color.



In **Y-Frame Classification** the menu includes a *Classification option* which displays a *verbal description* of the cluster, as well as predictive statistics for categorical dependent variable.



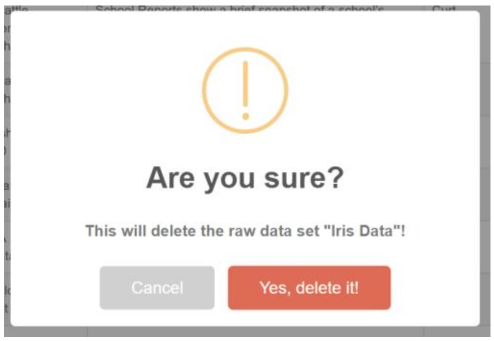


Deleting a Data Set or Model

To delete a model, or an entire data set, click on the trash can under "Delete Model" in the "**View and Manage Team Data**" tab.

Model Name	Description	Created By	Created Date	Delete Model	Rebuild Model	Analyze Model
Basice Consumer Loan Default Model 1.0	5 Independent Variables (proportionate scale) vs. 3 Dependent Variables	Curt Clausen	April 10, 2019 4:15 PM			

A failsafe pop-up will appear asking you to confirm.



Only the *[Creator of a data set or a model]* is allowed to delete their data sets or models.

Trouble Shooting

Server Error….the User is NULL. If you get this error message when uploading a file, it may be caused by a time-out in your browser connection to the Polynary service. Try signing out, clearing your cache (In Chrome, IE & Mozilla the keyboard shortcut is Shift+Ctrl+delete) and then logging back into your account.

Server Error in '/' Application.

The User is NULL. Please clear your cookies and try again. Specifically delete cookies for 'login.microsoftonline.com'. See this GitHub issue for more details: https://github.com/Azure-Samples/active-directory-b2c-dotnet-webapp-and-webapi/issues/9

Description: An unhandled exception occurred during the execution of the current web request. Please review the stack trace for more information about the error and where it originated in the code.

Exception Details: System.Exception: The User is NULL. Please clear your cookies and try again. Specifically delete cookies for 'login.microsoftonline.com'. See this GitHub issue for more details: https://github.com/Azure-Samples/active-directory-b2c-dotnet-webapp-and-webapi/issues/9

Source Error:

```
An unhandled exception was generated during the execution of the current web request. Information regarding the origin and location of the exception can be identified using the exception stack trace below.
```

Stack Trace:

```
[Exception: The User is NULL.  Please clear your cookies and try again.  Specifically delete cookies for 'login.microsoftonline.com'.  See this GitHub issue for more details: https://github.com
   PolynarySelfService.Helpers.<GetAccessTokenAsync>d__16.MoveNext() in C:\Users\erOxB\VSTF_Source\PolynaryService\PolynarySelfService\Helpers\PolynaryUser.cs:51
```

Support

For additional support please email: [info@polynary.com]. Our goal is to respond with 24 hours.

Glossary

Measurement Terms

Continuous variable: A continuous variable is a concept gauged on a numeric scale whose values reflect the amount, extent, or degree of something. Such scales have *directionality*; larger values mean 'more', and the distance between values has meaning. For instance, increasing weight is associated with being heavier, and a difference of five pounds between two people means the same thing regardless of their weights. Ratio and interval level scales are types of continuous variables.

Ordinal variable: An ordinal variable, usually coded as consecutive integers, also has directionality where larger values mean 'more' but unlike continuous variables the distance between its values has no meaning. For example, coding responses from a survey item: 1=not satisfied, 2= quite satisfied, 3=very satisfied we see that larger numbers are associated with greater satisfaction, but the distance between these responses are not assumed to be the same. Continuous variables whose values are replaced by their percentile rank are ordinal variables. As the number of response choices increases (say 5 or more) ordinal variables are often treated, for analytic purposes, as continuous variables.

Categorical variable: Also called nominal variables, a categorical variable has qualitative levels whose coding into numbers is arbitrary. For example, "Gender" coded: 1=Male, 2=Female could just as well be coded: 1=Female, 2=Male. Number values only serve to mark distinctions. Ordinal variables with less than 5 levels can be treated as categorical variables for analytic purposes.

RecordNames: The current program allows a text case label as the first column in the Xcel data input file. For instance, each row (case) may contain the name of a person. Certain options allow the user to identify the case labels associated with objects with similar properties.

Composite Variable: A composite variable arises when the values of multiple, closely related ordinal or continuous items are summed or averaged to produce a more reliable measure of the same underlying concept. For example, an overall Depression score might be defined as the average of 10 different assessment questions—each measured on a 5-pt scale.

Abstract Variables: It is not uncommon for people to express quantitative notions that cannot be assessed through the same kind of ruler. Unlike a composite variable, an abstraction is a concept made up of conceptually distinct notions constructed from the values of measured or less abstract notions. For instance, the overall assessment of a *waitperson's suitability* is an abstraction that might include a joint consideration of their ability to get along with other staff, have good customer relations, and cooperate with management expectations. Abstractions are multi-dimensional constructs channeled into a single continuous variable (such as suitability) through a joint consideration of its multiple dimensional values.

Abstract variables serve as a useful alternative to factor analysis. The current program allows the building of abstractions, bottom-up, from continuous variables into increasingly more abstract notions. This technique mirrors how people reduce hyper-dimensional problems into high-dimensional ones that respect the limits of working memory. Many quantitative problems of real-world complexity require the use of abstractions in order reflect the terms that we naturally use to think and talk about them.

Conceptual variable: A conceptual variable is a continuous variable whose values have been re-scaled to a relative value that falls between zero and one with larger values mean 'more' of that concept. This transformation frees a measure from arbitrary units of measure. For instance, a conceptual variable value of 0.5 represents a value that is 50% of the way along an interval of scale (***percent of scale***), or as ranked data, the 50^{th} ***percentile rank.***

Conceptual variables, whether understood in the context of percent of scale or the ranked sense, are critical to language. It is only in the context of a conceptual variable that adverbs, like "very" in the statement "John is very tall" makes sense. Polynary processes abstract and continuous variables as conceptual variables.

Analytic Terms

X-Frame: In Polynary, every analytic problem has an X-frame. An X-frame refers to the set of two or more conceptual variables used to describe, classify, or explain something. The 0-1 continuous scales of N independent (**input**) variables define a data space of an N-dimensional unit cube. Independent variables are also called *X-variables*. Polynary graphs show the partitions of the X-frame space.

Y-Frame: In Polynary, explanation problems also have a Y-frame of one or more conceptual variables that define what is "being explained". Just like X-Frames, the conceptual variables constitute a unit cube made up of dependent (**output**) variables. Dependent variables are also called *Y-variables*.

Y-Category: A Y-category is used to refer to a categorical variable used as a dependent variable. A Y-category is used in classification problems.

Prefilters: A Prefilter allows the user to display and perform a separate analysis for each level of a categorical/ordinal variable. This is important because patterns in data may differ across objects belonging to different categories. Prefilters are defined in the Excel input file.

Polynary model: Currently, Polynary models center around three frequently encountered analysis types: Description, Classification, and Explanation.

- *Description* is about characterizing collections of objects and is closely related to the tasks of identifying similar cases (matching) and cluster analysis. Description problems have an X-frame (and possibly a Prefilter) but no dependent (outcome) variable.

- *Classification* centers around distinguishing objects belonging to different categories. Analogous to discriminant analysis, classification models have a Y-category and an X-frame (and possibly a prefilter).

- *Explanation* is about accounting for something in terms of something else. Analogous to regression, explanations have a Y-frame (being explained) and an X-frame (doing the explaining). This problem type may also have a prefilter as an explanation may be different for objects belonging to different categories.

Methodology Terms

Polynary strings: The polynary representation is like a generalization of binary. Instead of just zeros and ones, a polynary string is composed of multiple symbols—one for each of the N conceptual dimensions plus a slack variable "v" that plays the same placeholding role as the zero in a binary string.

The string representation begins by transforming N Cartesian coordinates into (N+1) relative coordinates (an N-simplex). The values of the resulting relative coordinates have the properties of a set of proportions; they are all non-negative and together sum to a value of one. The relative magnitude of these proportions reflects the relative noteworthiness of each dimensional value. There is a one-to-one correspondence between a point in the N-dimensional Cartesian cube and a set of relative coordinates.

The polynary algorithm digitizes the values of a set of relative coordinates into a polynary string. The numeric precision of this re-representation converges quickly with increasing string length. The limited precision of measurement data puts one constraint of the length of strings needed. In data analysis the representation task is not about reproducing numeric precision of an individual object (point). Instead, it is about identifying the patterns and relationships formed over collections of points. These features are generalizations that occur over regions of space that may contain many objects. The length of polynary strings needed in analyses are those adequate to denote these regions. This length is considerably shorter than those needed to represent individual objects.

Geometrically, a polynary string denotes a chunk of N-dimensional Cartesian space called a basin. The longer the string, the smaller the chunk. The spelling of the string points to its location within the N-dimensional space. And from the spelling of this string we can determine the spelling of all its neighboring chunks. These realizations lead to approaching data analysis problems through spatial reasoning.

Spatial Clusters: We can denote larger and more arbitrary regions of space by joining adjacent basins (chunks of space). We call a connected set of basins a *spatial cluster* and define it through a list of polynary strings. There is no reason to assume that the way polynary partitions an N-dimensional space neatly aligns with the patterns or relationships we seek to identify from the data. Spatial clusters denote the regions of space around which generalizations about patterns or relationships in data are reported.

Natural Language Descriptions: A list of polynary strings denoting a spatial cluster is highly abstract. A more natural and widely understood approach is to describe these regions in language. This is precisely what quantitative descriptions do. They connote the location, size, and shape of a sub-region of a conceptual space. We have implemented this translation tool to make the reporting of results more human-friendly.

- **Accuracy:** Accuracy defined as the percentage of overlap between a denoted region of space and that connoted by its description (based on survey data from 100 native speakers of English). While language descriptions are never perfect, accuracies around 90% are not uncommon.
- **%Comp** (%Comprehensiveness)**:** As dimensionality increases, complete descriptions become too long to easily remember. Descriptions are made *briefer* by dropping the adverb-adjective phrases that convey the least amount of information. *%Comp* is a slider option that allows the user to select the percentage of information to retain.
- **Core %cases:** As the size of a spatial cluster increases its description becomes less specific. A common strategy to make descriptions more specific is to account for just the more typical members. For example, "The majority of our customers…". This slider option allows the user to modify the descriptions by changing the percent of cluster cases it covers.

Graphing Terms

Graph Resolution: The *Depth slider* allows the user to change the level of detail presented in a graph. The graph resolution is equal to the length of the polynary strings displayed.

Zooming: Another way to change the level of detail presented is to zoom in on the desired region on the fractal graph. This is done by clicking on any of the gray polygons found at all scales of the graph. To un-zoom the user can (repeatedly) click on the gray polygon at the center of the graph until the full graph is restored.

Y-variable intervals: When the Y-frame consists of a single continuous variable its range is divided into equally spaced, color-coded intervals. Modify the number of intervals through the *Intervals slider*.

Team Data Library: Team data libraries are user or team of users' secure repository for data sets uploaded to Polynary and models created from those data sets. There are some public team data libraries, like "Public Sandbox," that are open to users from various organizations. In these public libraries, data is not private.

Creator of a data set or a model: For data sets, this is the logged in user who uploaded the data set. For Polynary models, this is the logged in user who created the model.