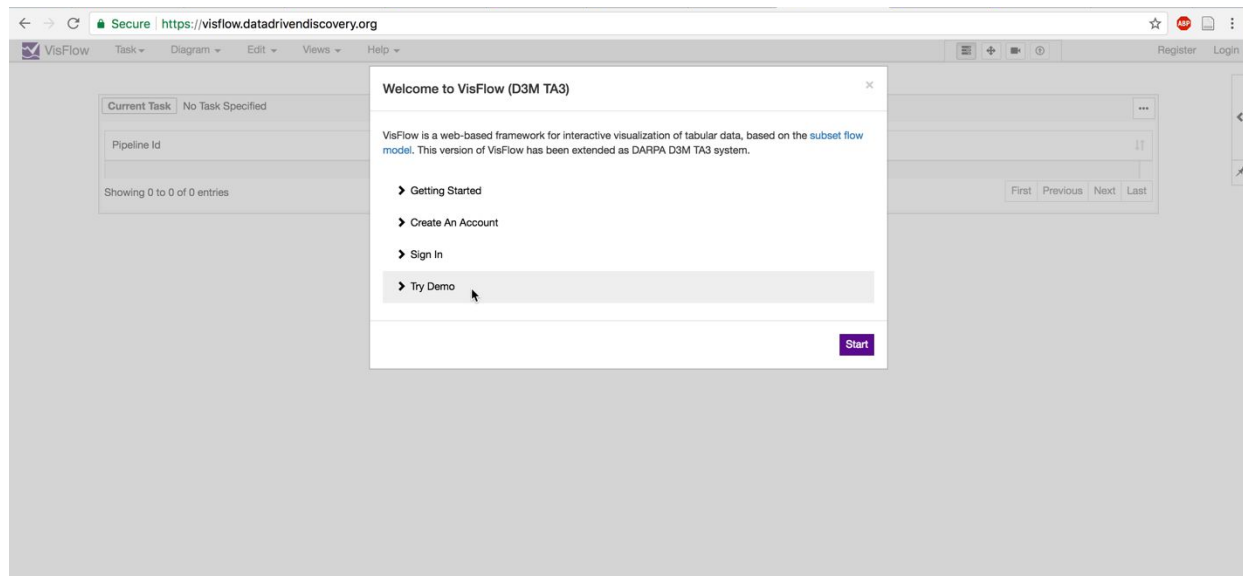


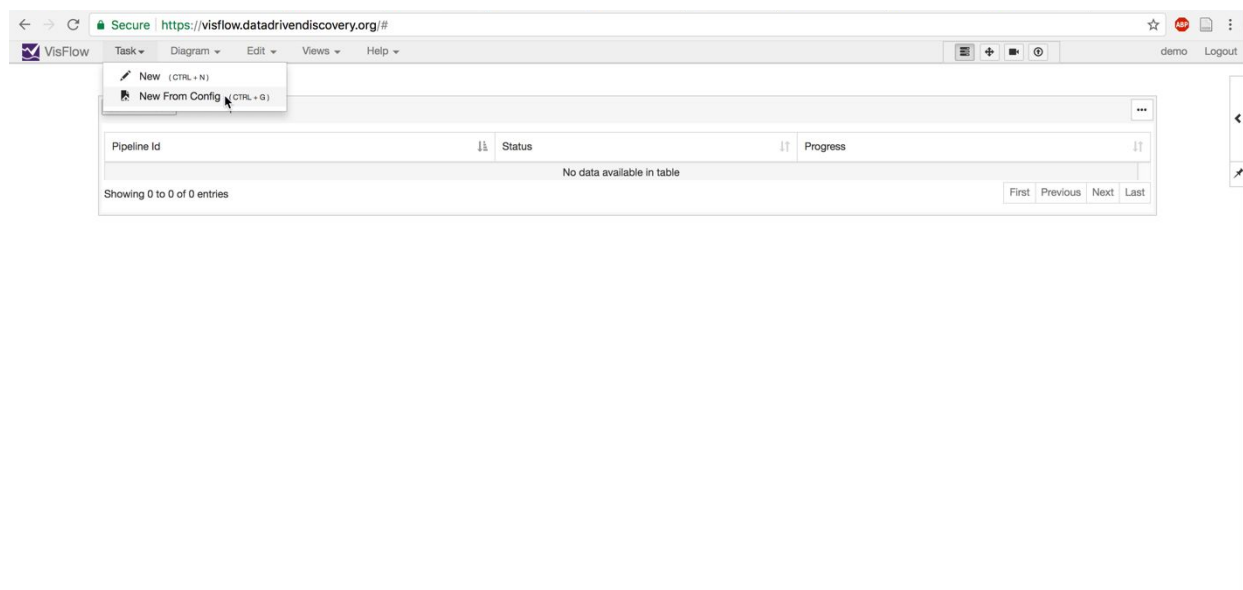
Visflow TA3 – Instructions

NEW YORK UNIVERSITY

1) Visit our TA3 at <https://visflow.datadrivendiscovery.org/> and click “Try Demo”:



2) To create a new Task from a configuration file, click “Task”, and then click “New From Config”:



3) The task will be launched and Visflow will display its execution progress:

Secure <https://visflow.data-drivendiscovery.org/#>

VisFlow Task Diagram Edit Views Help demo Logout

Current Task: o_185 classification/multiClass

Pipeline Id	Status	Progress	F1Macro
1eb41fc5-73d2-4a54-b0c8-6aedcfa263c9	ok	N/A	N/A
22afa7cc-4fe7-4279-b0a8-7c3dfbf5c15	ok	running	N/A
29278279-1235-4bcb-a928-3e98ef2e2c76	ok	N/A	N/A
3f1bbadb-4a98-4b99-8f52-8d1c8d114901	ok	completed	0.3310045301914215
5c09bb56-e5c7-4377-8207-df68b5bae92	ok	completed	0.5946746468544006

Showing 1 to 5 of 13 entries

First Previous Next Last

4) You can select a pipeline in the table and visualize its steps as a data flow:

Secure <https://visflow.data-drivendiscovery.org/#>

VisFlow Task Diagram Edit Views Help demo Logout

Current Task: o_185 classification/multiClass

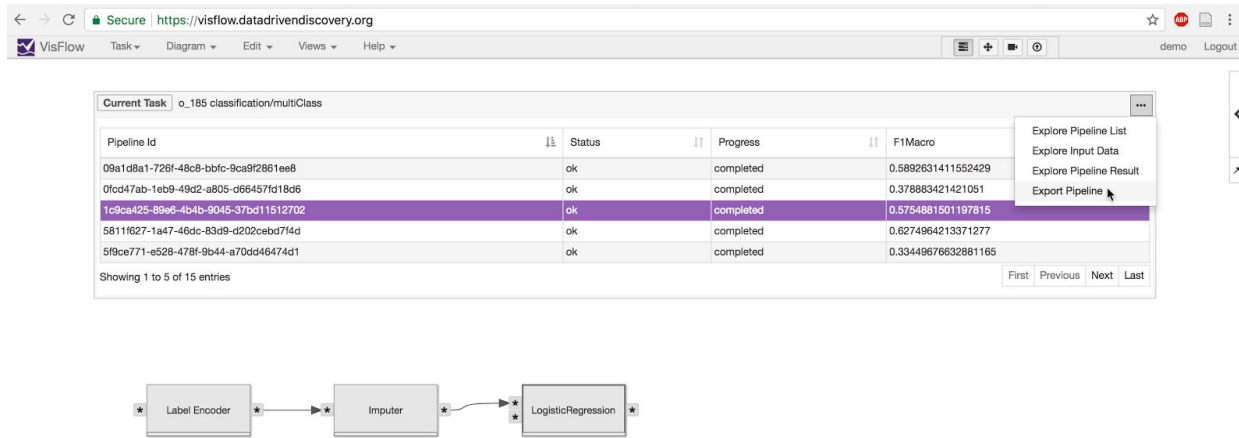
Pipeline Id	Status	Progress	F1Macro
09a1d8a1-726f-48c8-bbfc-9ca9f2861ee8	ok	completed	0.5892631411552429
0fcd47ab-1eb9-49d2-a805-d66457fd18d6	ok	completed	0.378883421421051
1c9ca425-89e6-4b4b-9045-37bd11512702	ok	completed	0.5754881501197815
5811f627-1a47-46dc-83d9-d202cebd7f4d	ok	completed	0.6274964213371277
5f9ce771-e528-478f-9b44-a70d3d46474d1	ok	completed	0.33449676632881165

Showing 1 to 5 of 15 entries

First Previous Next Last



5) Export the selected pipeline to the location specified by the configuration file by clicking “...” and “Export pipeline”:



The screenshot shows the VisFlow web interface at <https://visflow.data-drivendiscovery.org>. The current task is "o_185 classification/multiClass". A table lists several pipelines, with the third one selected. A dropdown menu is open, showing options: "Explore Pipeline List", "Explore Input Data", "Explore Pipeline Result", and "Export Pipeline". Below the table, a workflow diagram shows a sequence of steps: "Label Encoder" → "Imputer" → "LogisticRegression".

Pipeline Id	Status	Progress	F1Macro
09a1d8a1-726f-48c8-bbfc-9ca9f2861ee8	ok	completed	0.5892631411552429
0fc047ab-1eb9-49d2-a805-d66457fd18d6	ok	completed	0.378883421421051
1c9ca425-89e6-4b4b-9045-37bd11512702	ok	completed	0.5754881501197815
5811f627-1a47-46dc-83d9-d202cebd7f4d	ok	completed	0.6274964213371277
5f9ce771-e528-478f-9b44-a70dd46474d1	ok	completed	0.33449676632881165

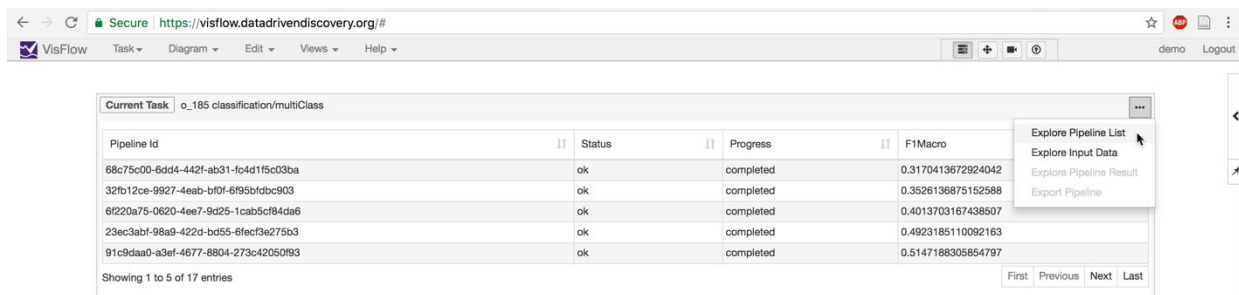
Showing 1 to 5 of 15 entries

First Previous Next Last

```

graph LR
    A[Label Encoder] --> B[Imputer]
    B --> C[LogisticRegression]
  
```

6) You can also explore the pipeline results using Visflow’s visualization capabilities. First click “Explore Pipeline List”:



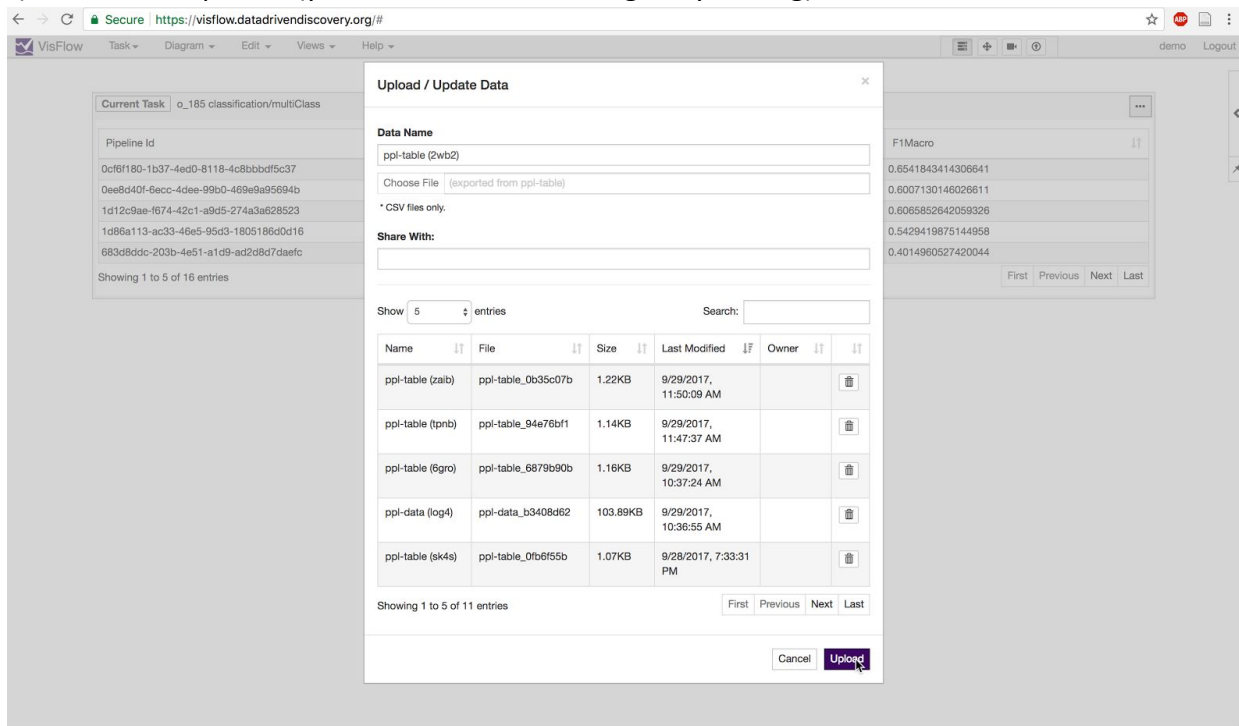
The screenshot shows the VisFlow web interface at <https://visflow.data-drivendiscovery.org/#>. The current task is "o_185 classification/multiClass". A table lists several pipelines. A dropdown menu is open, showing options: "Explore Pipeline List", "Explore Input Data", "Explore Pipeline Result", and "Export Pipeline". The "Explore Pipeline List" option is highlighted.

Pipeline Id	Status	Progress	F1Macro
68c75c00-6dd4-442f-ab31-fc4d1f5c03ba	ok	completed	0.3170413672924042
32fb12ce-9927-4eab-bf0f-6f95bfd9c903	ok	completed	0.3526136875152588
6f220a75-0620-4ee7-9d25-1cab5cf84da6	ok	completed	0.4013703167438507
23ec3abf-98a9-422d-bd55-fecf3a275b3	ok	completed	0.4923185110092163
91c9daa0-a3ef-4677-8804-273c42050f93	ok	completed	0.5147188305854797

Showing 1 to 5 of 17 entries

First Previous Next Last

7) Then click “upload” (you do not need to change any setting):



Upload / Update Data

Data Name: ppl-table (2vb2)

Choose File: (exported from ppl-table)

* CSV files only.

Share With:

Show 5 entries Search:

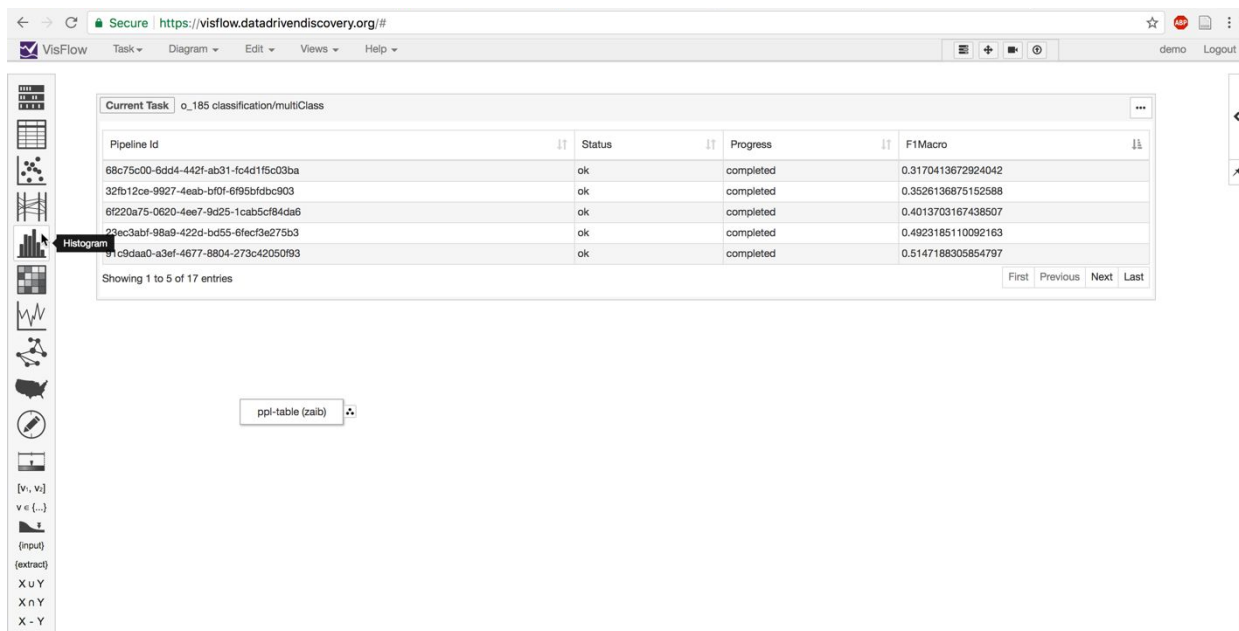
Name	File	Size	Last Modified	Owner	
ppl-table (zaib)	ppl-table_0b35c07b	1.22KB	9/29/2017, 11:50:09 AM		
ppl-table (tpnb)	ppl-table_94e76bf1	1.14KB	9/29/2017, 11:47:37 AM		
ppl-table (6gro)	ppl-table_6879b90b	1.16KB	9/29/2017, 10:37:24 AM		
ppl-data (log4)	ppl-data_b3408d62	103.89KB	9/29/2017, 10:36:55 AM		
ppl-table (sk4s)	ppl-table_0fb6f55b	1.07KB	9/28/2017, 7:33:31 PM		

Showing 1 to 5 of 11 entries

First Previous Next Last

Cancel Upload

8) A Data Source will appear on your screen (ppl-table):



Current Task o_185 classification/multiClass

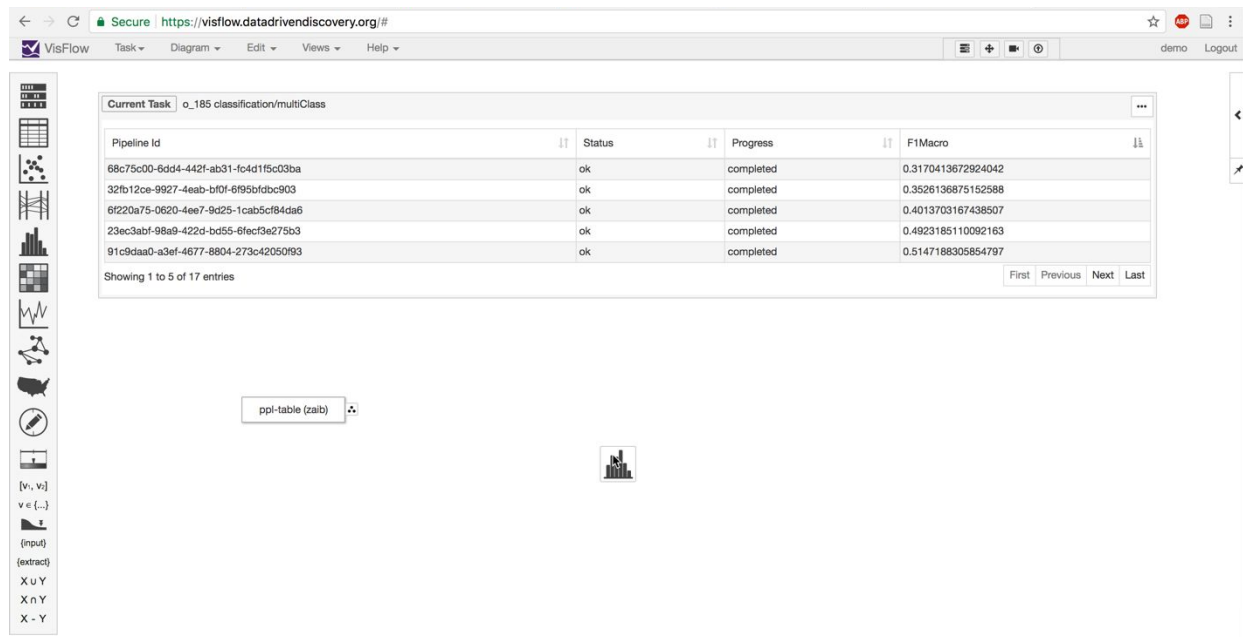
Pipeline Id	Status	Progress	F1Macro
68c75c00-6dd4-442f-ab31-fc4d1f5c03ba	ok	completed	0.3170413672924042
32fb12ce-9927-4eab-bf01-6f95bfdbc903	ok	completed	0.3526136875152588
6f220a75-0620-4ee7-9d25-1cab5cf84da6	ok	completed	0.4013703167438507
22ec3abf-98a9-422d-bd55-6fecf3e275b3	ok	completed	0.4923185110092163
31c8daa0-a3ef-4677-8804-273c42050f93	ok	completed	0.5147188305854797

Showing 1 to 5 of 17 entries

First Previous Next Last

ppl-table (zaib)

9) Next, drag the histogram to the workspace:

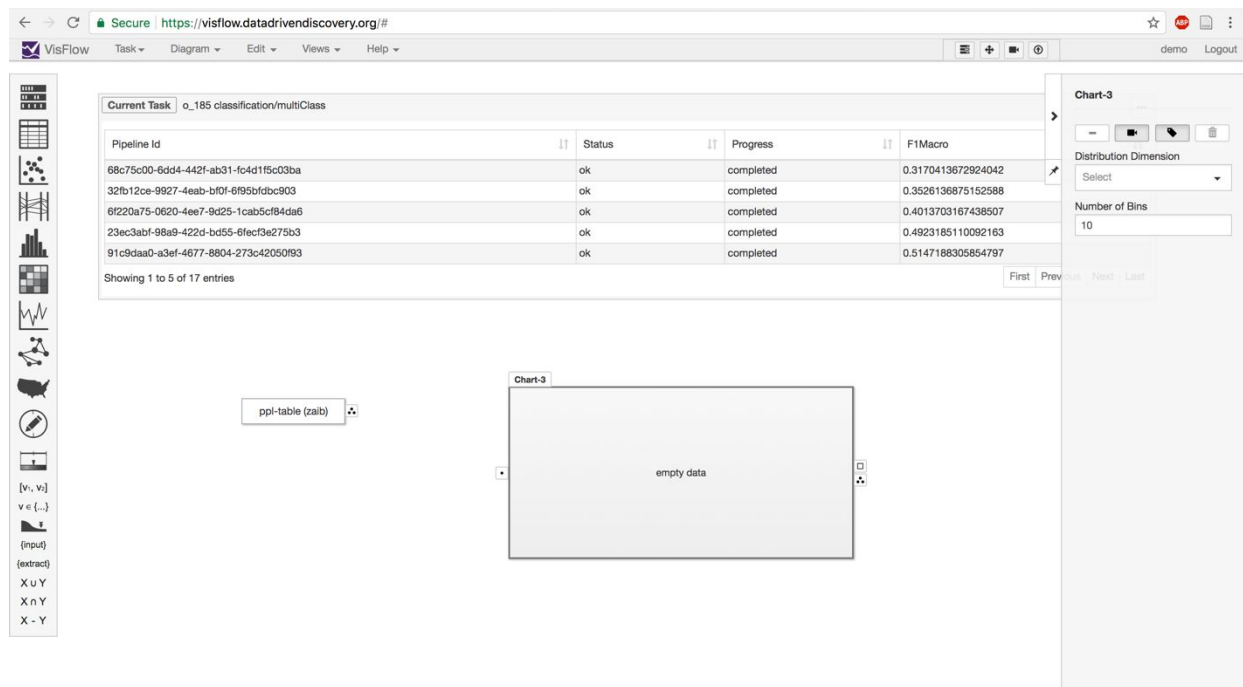


The screenshot shows the VisFlow interface with a table of pipeline results. The table has columns for Pipeline Id, Status, Progress, and F1Macro. The data is as follows:

Pipeline Id	Status	Progress	F1Macro
68c75c00-6dd4-442f-ab31-fc4d1f5c03ba	ok	completed	0.3170413672924042
32fb12ce-9927-4eab-bf0f-6f95bfdbc903	ok	completed	0.3526136875152588
6f220a75-0620-4ee7-9d25-1cab5cf84da6	ok	completed	0.4013703167438507
23ec3abf-98a9-422d-bd55-6fecf3e275b3	ok	completed	0.4923185110092163
91c9daa0-a3ef-4677-8804-273c42050f93	ok	completed	0.5147188305854797

Showing 1 to 5 of 17 entries

In the workspace, there is a small histogram icon and a label 'ppi-table (zaib)'.

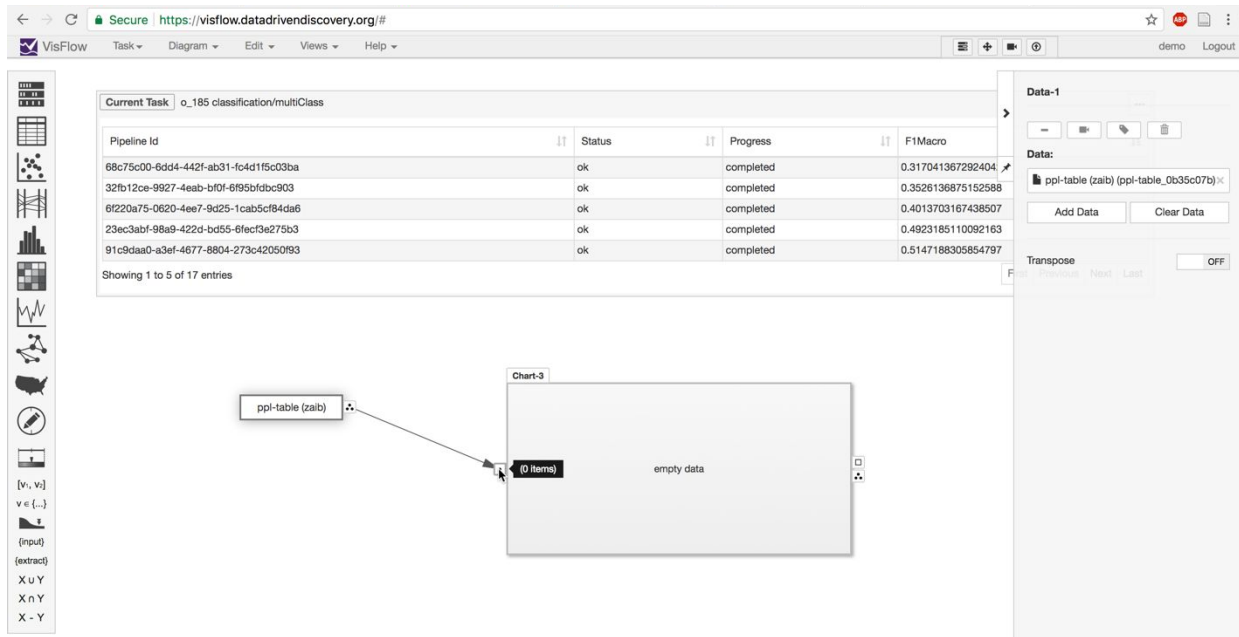


The screenshot shows the VisFlow interface with the histogram chart added to the workspace. The chart is labeled 'Chart-3' and contains the text 'empty data'. The workspace also contains the 'ppi-table (zaib)' label.

The chart configuration panel on the right shows the following settings:

- Chart-3
- Distribution Dimension: Select
- Number of Bins: 10

10) Finally, connect the output port of “ppl-table” to the input of your histogram (drag the data source port output to the input of the chart):

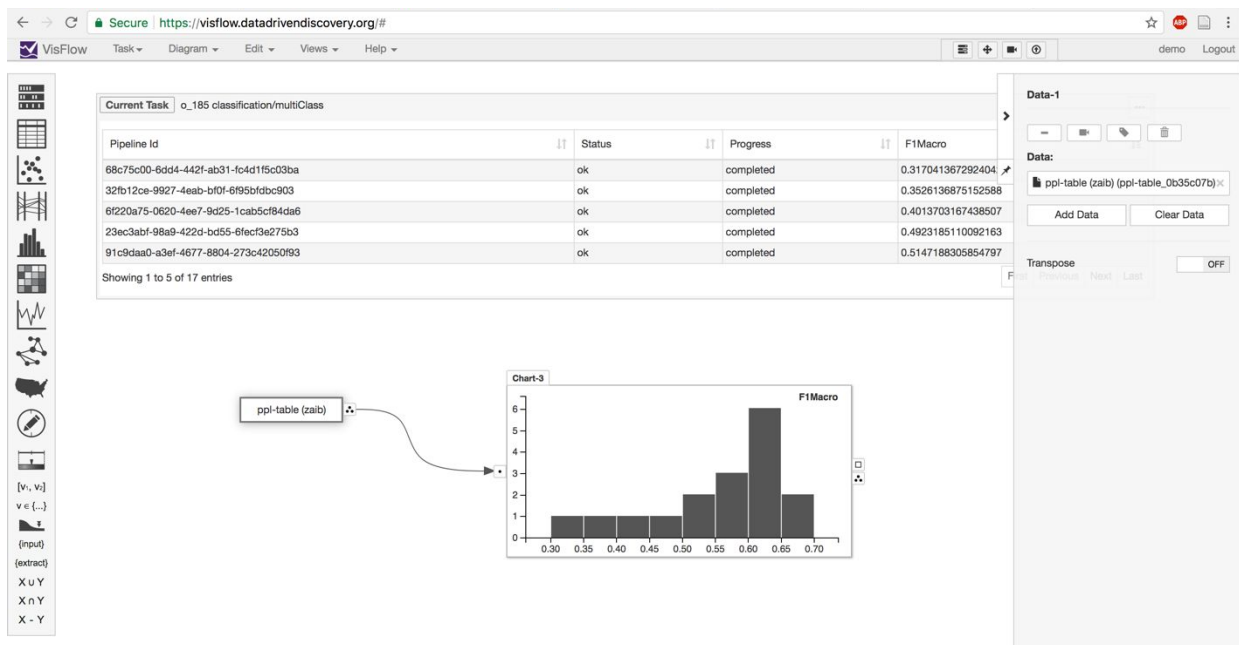


The screenshot shows the VisFlow interface with a task table and a disconnected chart. The task table is titled "Current Task | o_185 classification/multiClass" and contains 17 entries. The chart, labeled "Chart-3", is currently empty and disconnected from the "ppl-table (zaib)" data source.

Pipeline Id	Status	Progress	F1Macro
68c75c00-6dd4-442f-ab31-fc4d1f5c03ba	ok	completed	0.317041367292404
32fb12ce-9927-4eab-bf0f-6f95bfdc903	ok	completed	0.3526136875152588
6f220a75-0620-4ee7-9d25-1cab5cf84da6	ok	completed	0.4013703167438507
23ec3abf-98a9-422d-bd55-6fecf3e275b3	ok	completed	0.4923185110092163
91c9daa0-a3ef-4677-8804-273c42050f93	ok	completed	0.5147188305854797

Showing 1 to 5 of 17 entries

11) You should see a histogram of the quality measures from the created task:

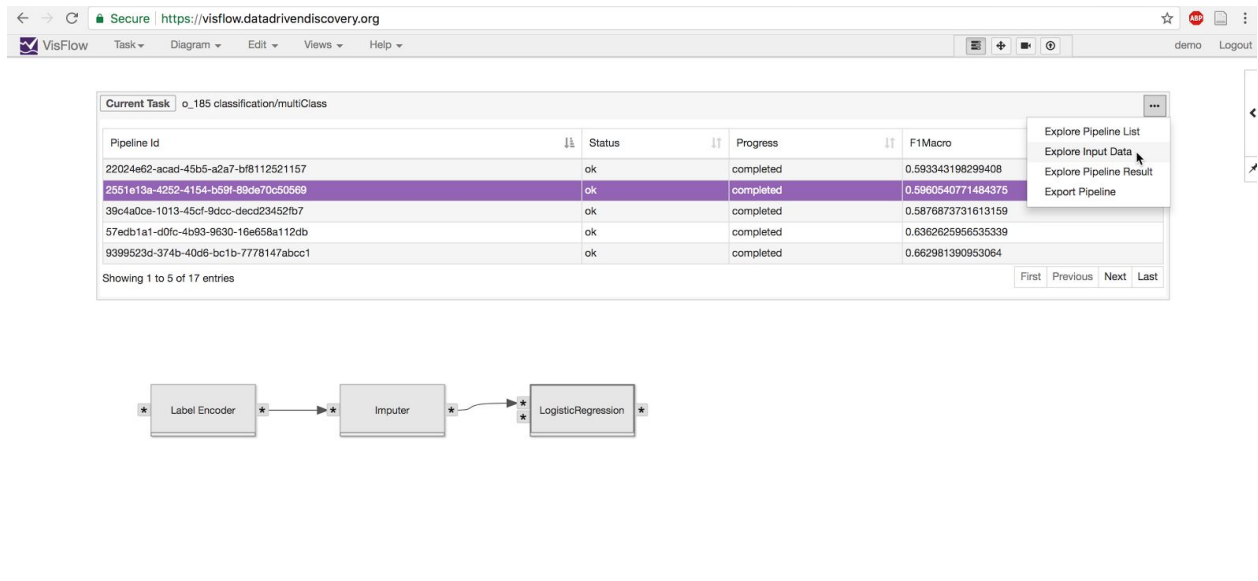


The screenshot shows the VisFlow interface with the same task table as before. The chart, labeled "Chart-3", now displays a histogram of the F1Macro values from the task table. The x-axis is labeled "F1Macro" and ranges from 0.30 to 0.70. The y-axis represents frequency, ranging from 0 to 6. The histogram shows a distribution of F1Macro values, with a peak around 0.60.

F1Macro Range	Frequency
0.30 - 0.35	1
0.35 - 0.40	1
0.40 - 0.45	1
0.45 - 0.50	1
0.50 - 0.55	2
0.55 - 0.60	3
0.60 - 0.65	6
0.65 - 0.70	2

OPTIONAL:

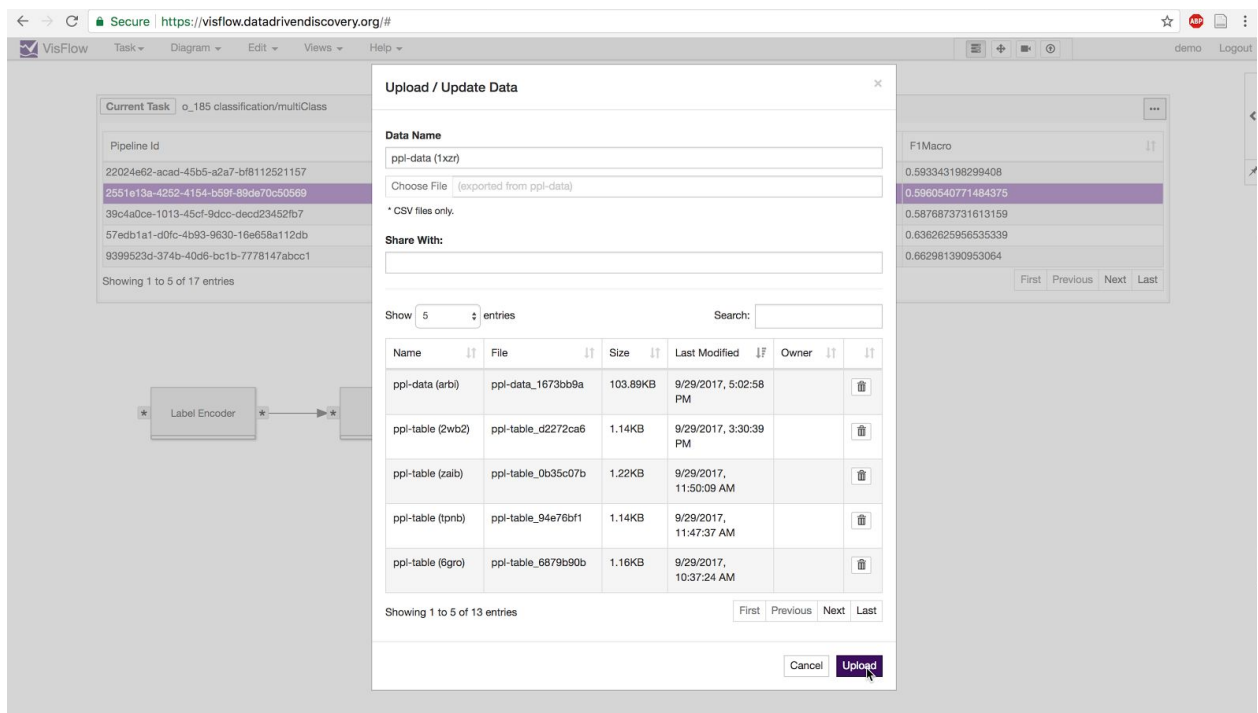
12) You can also visualize the training dataset using Visflow. First export the data to visflow, by clicking “Explore Input Data”:



The screenshot shows the VisFlow web interface at <https://visflow.data.drivendiscovery.org>. The 'Current Task' is 'o_185 classification/multiClass'. A table lists several pipelines, with the second one highlighted. A dropdown menu is open for the highlighted pipeline, showing options: 'Explore Pipeline List', 'Explore Input Data', 'Explore Pipeline Result', and 'Export Pipeline'. Below the table, a workflow diagram shows three steps: 'Label Encoder' → 'Imputer' → 'LogisticRegression'.

Pipeline Id	Status	Progress	F1Macro
22024e62-acad-45b5-a2a7-bf8112521157	ok	completed	0.593343198299408
2551e13a-4252-4154-b59f-89de70c50569	ok	completed	0.5960540771484375
39c4a0ce-1013-45cf-9dcc-decd23452fb7	ok	completed	0.5876873731613159
57ecb1a1-d0fc-4b93-9630-16e658a112db	ok	completed	0.6362625956535339
9399523d-374b-40d6-bc1b-7778147abcc1	ok	completed	0.662981390953064

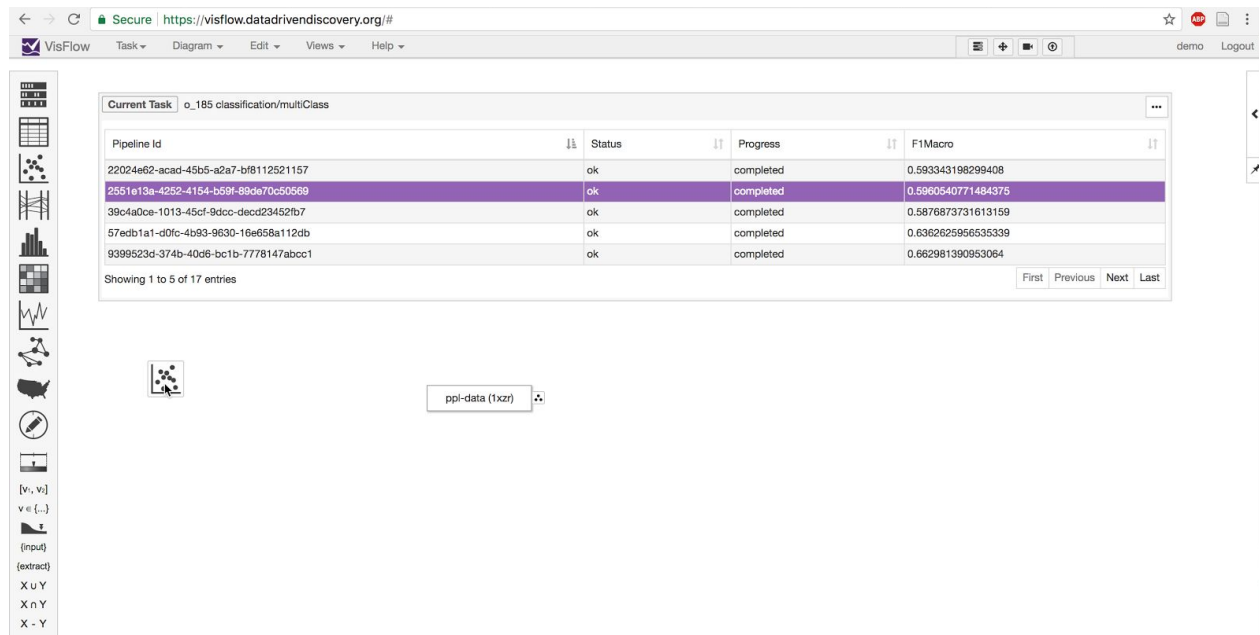
13) Then click upload:



The screenshot shows the 'Upload / Update Data' dialog box in the VisFlow web interface. The dialog has fields for 'Data Name' (ppl-data (1xzt)) and 'Choose File' (exported from ppl-data). Below these, there is a 'Share With' field and a 'Show' dropdown set to '5 entries'. A table lists the uploaded files with columns: Name, File, Size, Last Modified, Owner, and an action icon. The 'Upload' button is highlighted.

Name	File	Size	Last Modified	Owner	
ppl-data (arb)	ppl-data_1673bb9a	103.89KB	9/29/2017, 5:02:58 PM		
ppl-table (2wb2)	ppl-table_c2272ca6	1.14KB	9/29/2017, 3:30:39 PM		
ppl-table (zaib)	ppl-table_0b35c07b	1.22KB	9/29/2017, 11:50:09 AM		
ppl-table (tpnb)	ppl-table_94e76bf1	1.14KB	9/29/2017, 11:47:37 AM		
ppl-table (6gro)	ppl-table_6679b90b	1.16KB	9/29/2017, 10:37:24 AM		

14) Drag a scatter plot to the workspace:

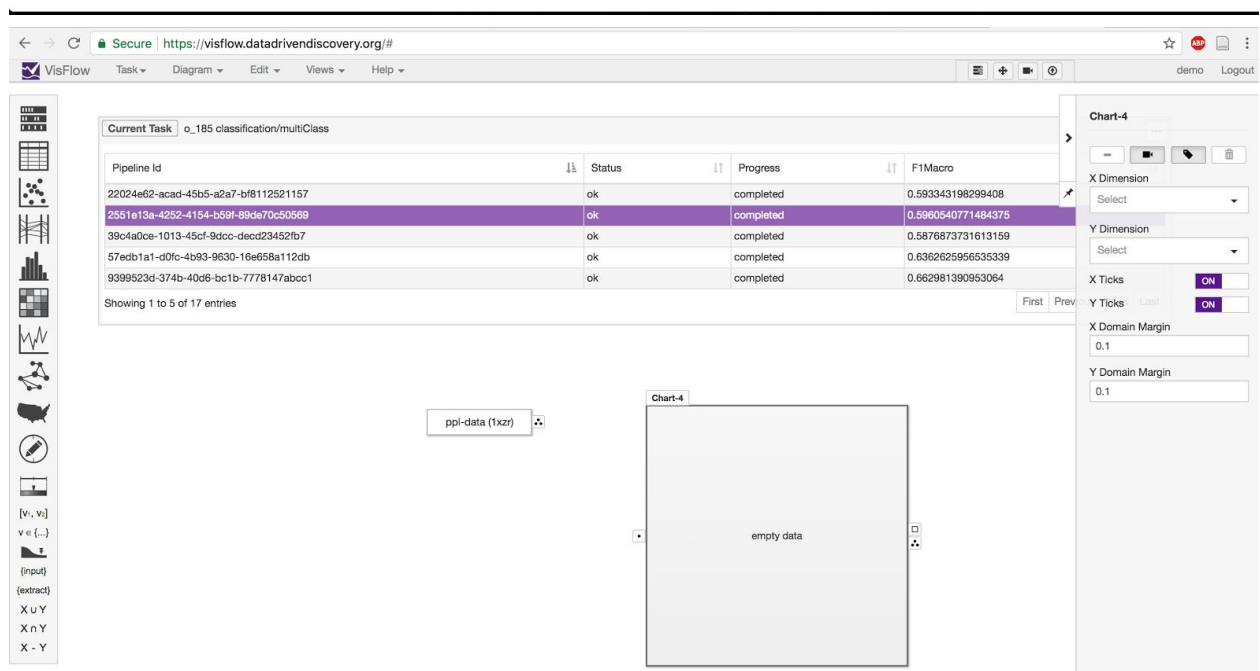


The screenshot shows the VisFlow interface with the 'Current Task' table and a workspace. The table lists pipeline IDs, status, progress, and F1Macro scores. The workspace contains a scatter plot icon and a data source labeled 'ppi-data (1x3)'.

Pipeline Id	Status	Progress	F1Macro
22024e62-acad-45b5-a2a7-bf8112521157	ok	completed	0.593343198299408
2551e13a-4252-4154-b59f-89de70c50569	ok	completed	0.5960540771484375
39c4a0ce-1013-45cf-9dcc-decd23452fb7	ok	completed	0.5876873731613159
57ed1a1-d0fc-4b93-9630-16e658a112db	ok	completed	0.6362625956535339
9399523d-374b-40d6-bc1b-7778147abcc1	ok	completed	0.662981390953064

Showing 1 to 5 of 17 entries

ppi-data (1x3)



The screenshot shows the VisFlow interface with the 'Current Task' table, a workspace, and a 'Chart-4' configuration panel. The table lists pipeline IDs, status, progress, and F1Macro scores. The workspace contains a scatter plot icon and a data source labeled 'ppi-data (1x3)'. The 'Chart-4' panel shows configuration options for X and Y dimensions, ticks, and margins.

Pipeline Id	Status	Progress	F1Macro
22024e62-acad-45b5-a2a7-bf8112521157	ok	completed	0.593343198299408
2551e13a-4252-4154-b59f-89de70c50569	ok	completed	0.5960540771484375
39c4a0ce-1013-45cf-9dcc-decd23452fb7	ok	completed	0.5876873731613159
57ed1a1-d0fc-4b93-9630-16e658a112db	ok	completed	0.6362625956535339
9399523d-374b-40d6-bc1b-7778147abcc1	ok	completed	0.662981390953064

Showing 1 to 5 of 17 entries

ppi-data (1x3)

Chart-4

X Dimension: Select

Y Dimension: Select

X Ticks: ON

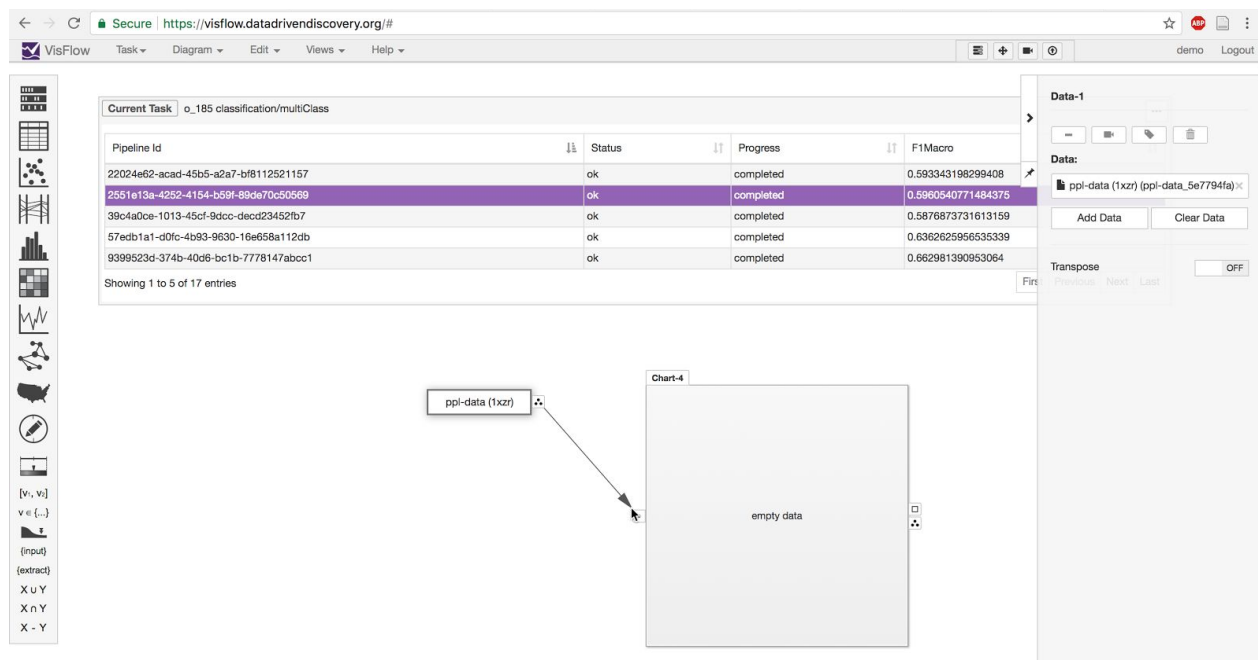
Y Ticks: ON

X Domain Margin: 0.1

Y Domain Margin: 0.1

empty data

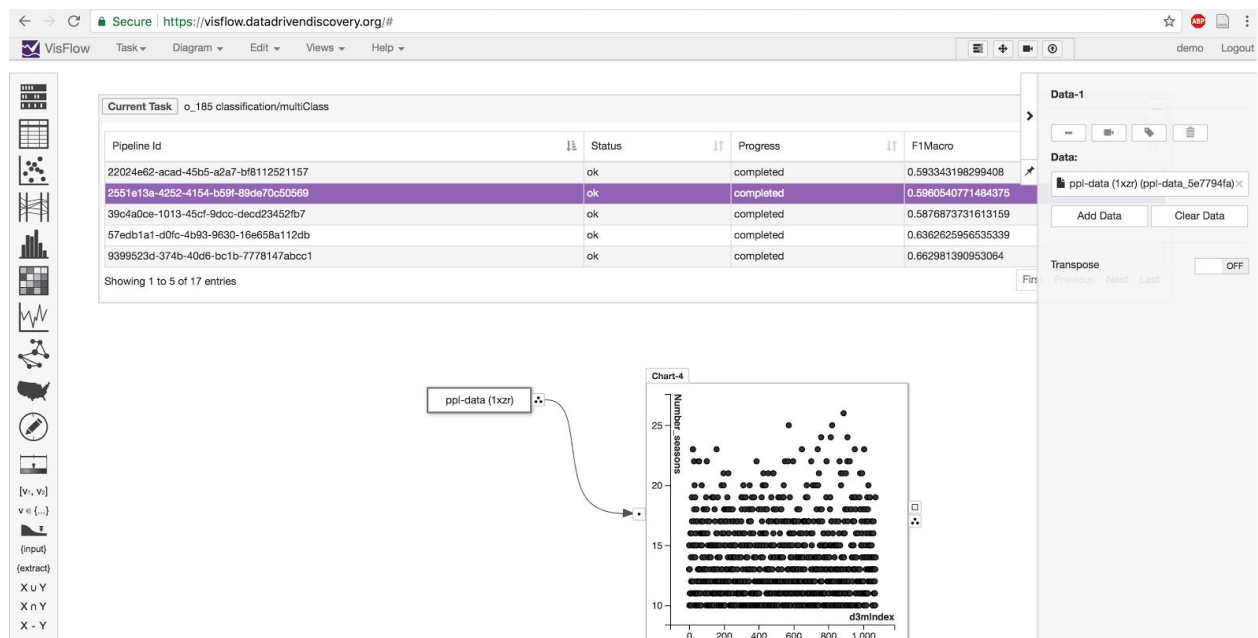
15) And connect the output of the data to the input of your scatter plot chart:



The screenshot shows the VisFlow interface with a pipeline table and a scatter plot chart. The pipeline table lists several pipelines, with the one having ID 2551e13a-4252-4154-b59f-89de70c50569 highlighted. The scatter plot chart, labeled Chart-4, is currently empty, showing the text "empty data". A connection arrow points from the "ppi-data (1x2r)" output of the pipeline to the input of the scatter plot chart.

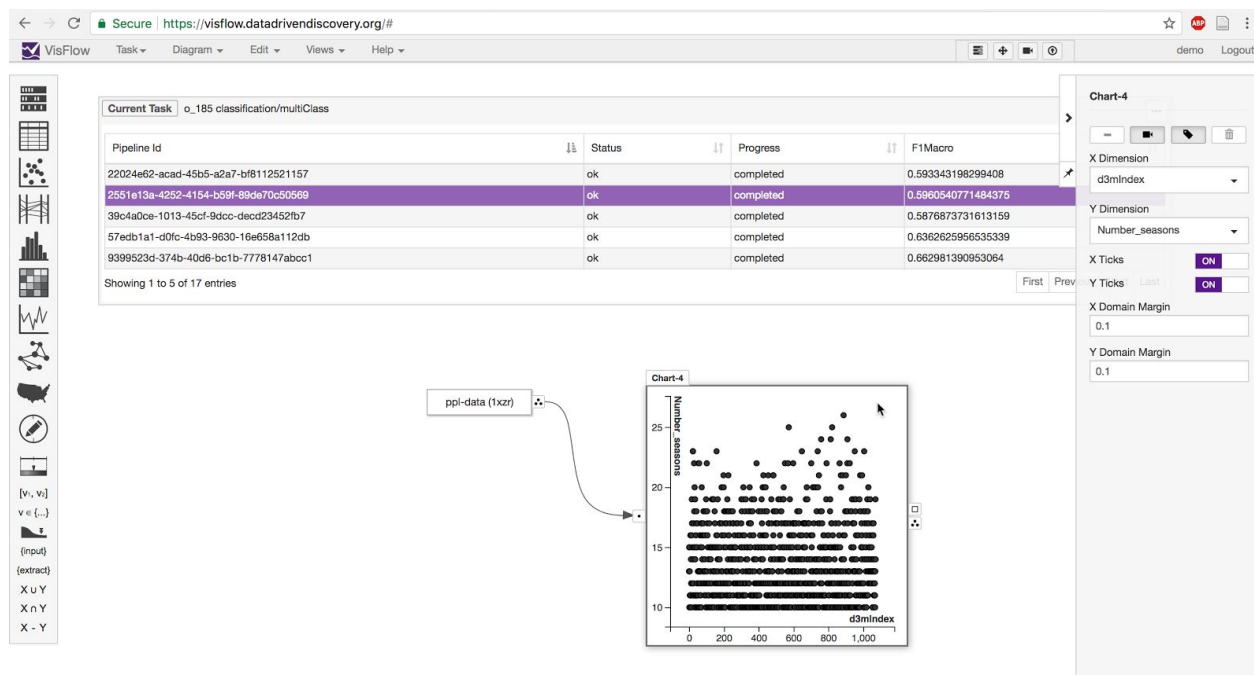
Pipeline Id	Status	Progress	F1Macro
22024e62-acad-45b5-a2a7-bf8112521157	ok	completed	0.593343198299408
2551e13a-4252-4154-b59f-89de70c50569	ok	completed	0.5960540771484375
39c4a0ce-1013-45cf-9dcc-decd23452fb7	ok	completed	0.5876873731613159
57edcb1a1-d0fc-4b93-9630-16e658a112db	ok	completed	0.6362625956535339
9399523d-374b-40d8-bc1b-7778147abcc1	ok	completed	0.662981390953064

16) You should see a scatter plot of two dimensions of the data:

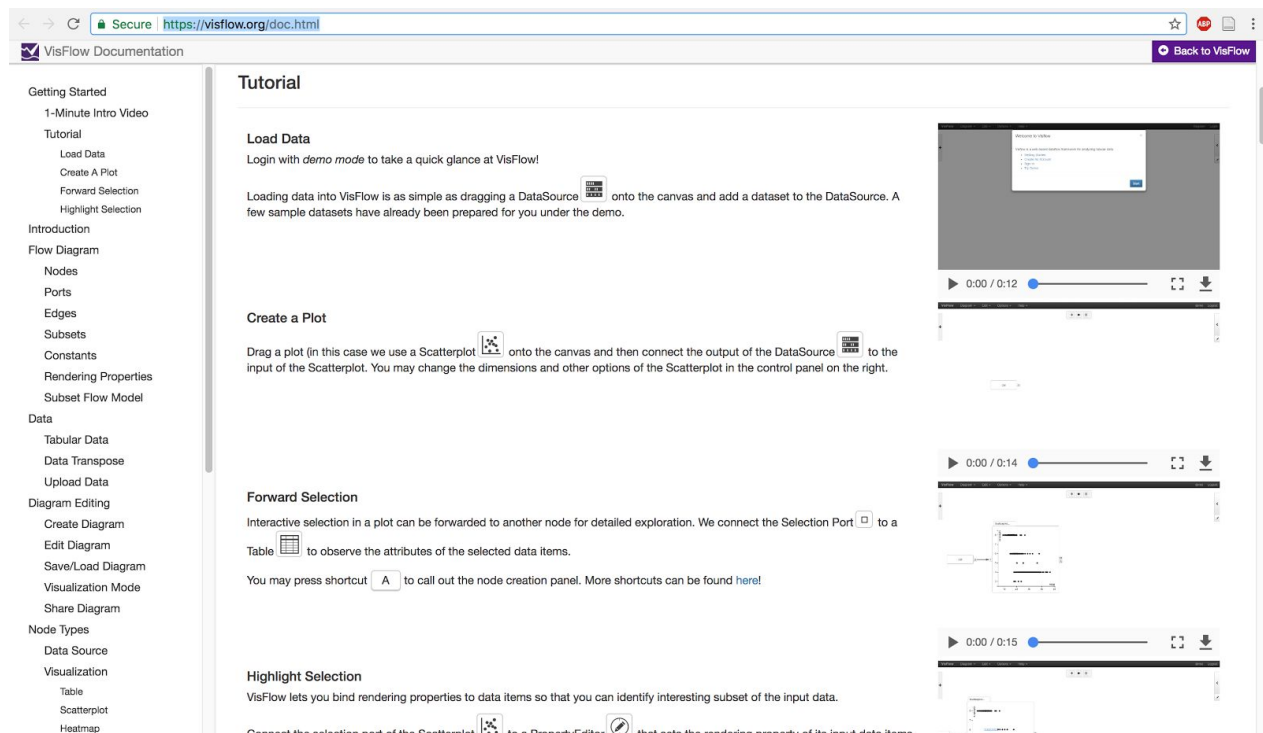


The screenshot shows the VisFlow interface with the same pipeline table as before. The scatter plot chart, labeled Chart-4, now displays a scatter plot of two dimensions of the data. The x-axis is labeled "d3mIndex" and ranges from 0 to 1,000. The y-axis is labeled "Number seasons" and ranges from 10 to 25. The data points are clustered, showing a positive correlation between the two dimensions.

17) Click the scatter plot, and you can change the axis dimensions:



18) If you want to learn more about visflow, visit our documentation at <https://visflow.org/doc.html>



The screenshot shows the VisFlow documentation page. The left sidebar contains a navigation menu with sections like 'Getting Started', 'Tutorial', 'Introduction', 'Flow Diagram', 'Nodes', 'Ports', 'Edges', 'Subsets', 'Constants', 'Rendering Properties', 'Subset Flow Model', 'Data', 'Diagram Editing', 'Create Diagram', 'Edit Diagram', 'Save/Load Diagram', 'Visualization Mode', 'Share Diagram', 'Node Types', 'Data Source', 'Visualization', 'Table', 'Scatterplot', and 'Heatmap'.

The main content area is titled 'Tutorial' and contains sections for 'Load Data', 'Create a Plot', 'Forward Selection', and 'Highlight Selection'. Each section includes text instructions and a small diagram or video player. The 'Load Data' section shows how to login with 'demo mode'. The 'Create a Plot' section shows how to drag a 'Scatterplot' node onto the canvas and connect it to a 'DataSource' node. The 'Forward Selection' section shows how to connect the 'Selection Port' of a plot to a 'Table' node. The 'Highlight Selection' section shows how to connect the 'selection port' of the 'Scatterplot' to a 'PropertyEditor' node.