R Script add column name

IMPORT DATASET

R Script add column name

In this session

- Import dataset from local CVS file to R data frame (R Studio)
- Add column name Using R code (R Studio)
- Import dataset from internet (R Studio)
- Import dataset from internet (Azure ML Studio)
- Add column name Using R code (Azure ML Studio)
- Data Visualization (Azure ML Studio)

R Script add column name

Import dataset from local CVS file to R data frame (R Studio)

adult100.csv

	A	B	C	D	E	F
1	age	workclass	marital-status	occupation	race	sex
2	39	State-gov	Never-married	Adm-clerical	White	Male
3	50	Self-emp-not-inc	Married-civ-spouse	Exec-managerial	White	Male
4	38	Private	Divorced	Handlers-cleaners	White	Male
5	53	Private	Married-civ-spouse	Handlers-cleaners	Black	Male
6	28	Private	Married-civ-spouse	Prof-specialty	Black	Female
7	37	Private	Married-civ-spouse	Exec-managerial	White	Female
8	49	Private	Married-spouse-absent	Other-service	Black	Female
9	52	Self-emp-not-inc	Married-civ-spouse	Exec-managerial	White	Male
10	31	Private	Never-married	Prof-specialty	White	Female
11	42	Private	Married-civ-spouse	Exec-managerial	White	Male

R Script add column name

Import dataset from local CVS file to R data frame (R Studio)

```
getwd()                                  # get working directory
setwd("c:/temp")                         # set working directory
list.files()                             # list file in current directory
d1 <- read.csv("adult100.csv", header = FALSE)
str(d1)                                  # show structure of d1
```

```
'data.frame':   99 obs. of  6 variables:
 $ age           : int  39 50 38 53 28 37 49 52 31 42 ...
 $ workclass     : Factor w/ 7 levels " ?"," Federal-gov",..: 7 6 4 4 4 4 4 6 4 4 ...
 $ marital-status: Factor w/ 6 levels " Divorced"," Married-AF-spouse",..: 5 3 1 3 3 3 4 3 5 3 ...
 $ occupation    : Factor w/ 13 levels " ?"," Adm-clerical",..: 2 4 6 6 9 4 8 4 9 4 ...
 $ race          : Factor w/ 5 levels " Amer-Indian-Eskimo",..: 5 5 5 3 3 5 3 5 5 5 ...
 $ sex           : Factor w/ 2 levels " Female"," Male": 2 2 2 2 1 1 1 2 1 2 ...
```

```
typeof(d1)                               # show data type of d1
class(d1)                                # show class of d1
dim(d1)                                  # show dimension of d1
```

R Script add column name

Import dataset from local CVS file to R data frame (R Studio)

```
ncol(d1)                    # number of columns
nrow(d1)                    # number of rows
d1                          # show all data rows
head(d1)                    # preview data in d1
head(d1, n = 10)            # show top 10 rows
tail(d1)                    # preview data in d1
tail(d1, n = 10)            # show bottom 10 rows
d1[1:3]                     # show column 1 to column 3
head(d1[1:3], n = 10)       # show top 10 rows column 1 to column 3
```

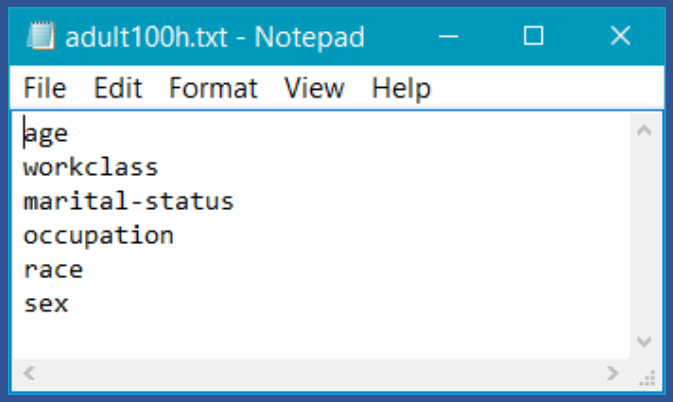
R Script add column name

Add column name Using R code (R Studio)

Adult100n.csv

	A	B	C	D	E	F
1	39	State-gov	Never-ma	Adm-cleri	White	Male
2	50	Self-emp-r	Married-c	Exec-mana	White	Male
3	38	Private	Divorced	Handlers-c	White	Male
4	53	Private	Married-c	Handlers-c	Black	Male
5	28	Private	Married-c	Prof-speci	Black	Female
6	37	Private	Married-c	Exec-mana	White	Female
7	49	Private	Married-s	Other-serv	Black	Female

Adult100h.txt



R Script add column name

Add column name Using R code (R Studio)

```
ls()                                    # print all object in workspace
rm(list=ls())                           # Clear R workspace
d1 <- read.csv("adult100n.csv", header = FALSE)  # import dataset without column
** typeof, structure, class, preview d1
d2 <- readLines("adult100h.txt")        # import column name
** typeof, structure, class, preview d2
colnames(d1) <- d2                      # update d1 column names
** preview d1
```

R Script add column name

Import dataset from internet (R Studio)

UCI Machine Learning Repository: Adult Data Set

home page

https://archive.ics.uci.edu/ml/datasets/adult

Description

https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.names

Data Set

http://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data

R Script add column name

Import dataset from internet (R Studio)

UCI

About Citation Policy Donate a Data Set Contact

Search

Repository Web Google

Machine Learning Repository

Center for Machine Learning and Intelligent Systems

View ALL Data Sets

Adult Data Set

Download: Data Folder, Data Set Description

Abstract: Predict whether income exceeds $50K/yr based on census data. Also known as "Census Income" dataset.

Data Set Characteristics:	Multivariate	**Number of Instances:**	48842	**Area:**	Social
Attribute Characteristics:	Categorical, Integer	**Number of Attributes:**	14	**Date Donated**	1996-05-01
Associated Tasks:	Classification	**Missing Values?**	Yes	**Number of Web Hits:**	899430

R Script add column name

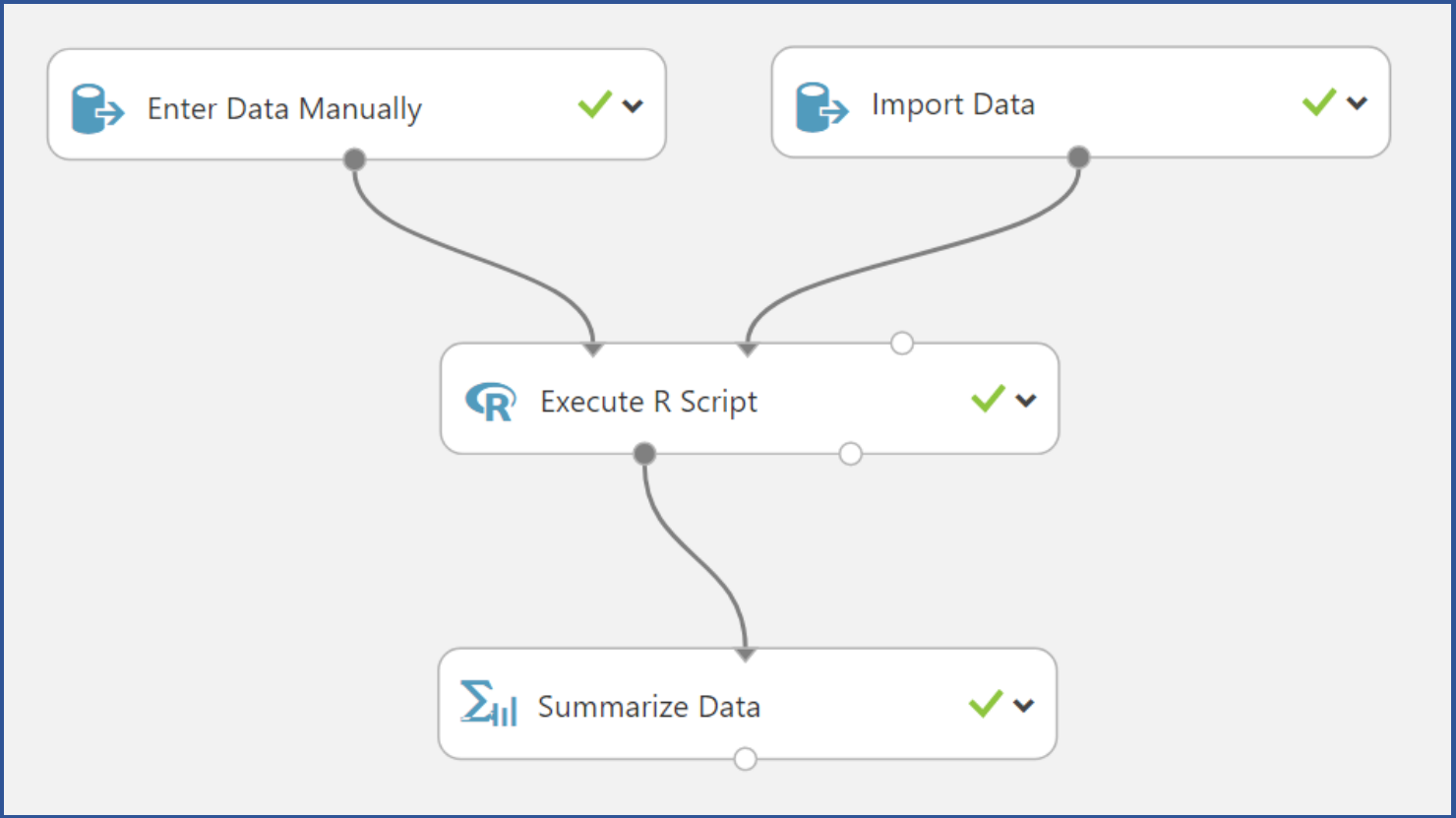
Import dataset from internet (R Studio)

```
rm(list=ls())   # Clear R workspace
u <- "http://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data"
d1 <- read.csv(url(u), header = FALSE)    # import dataset
View(d1)              # invoke spreadsheet-style data viewer on a matrix-like R object
d2 <- readLines("cencol.txt")    # import column name
colnames(d1) <- d2     # update d1 column names
```

R Script add column name

Import dataset from internet (Azure ML Studio)

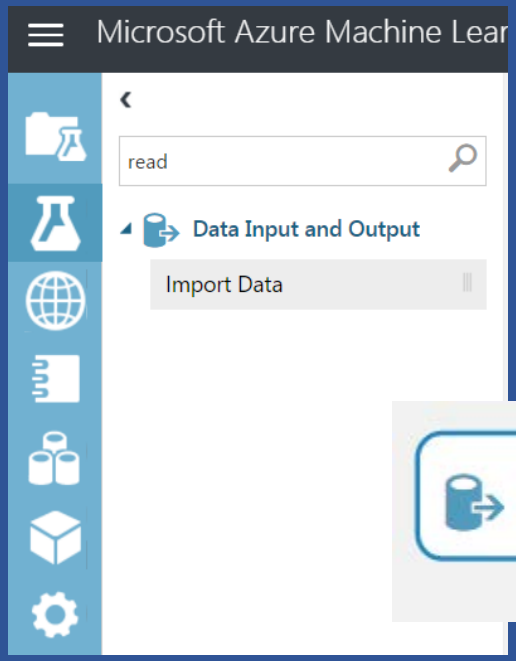
This experiment on completed



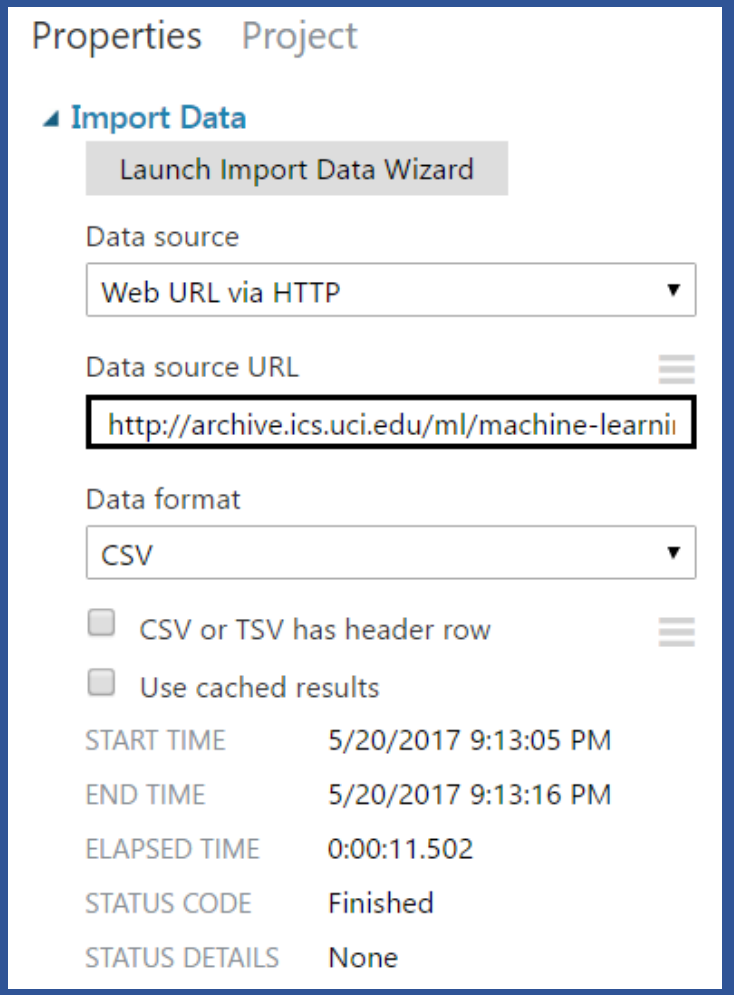
R Script add column name

Import dataset from internet (Azure ML Studio)

- Open Microsoft Azure Machine Learning Studio
- Create New blank experiment name = R add col name
- Click Data Input and Output
- Drag & drop Import Data
- Set properties



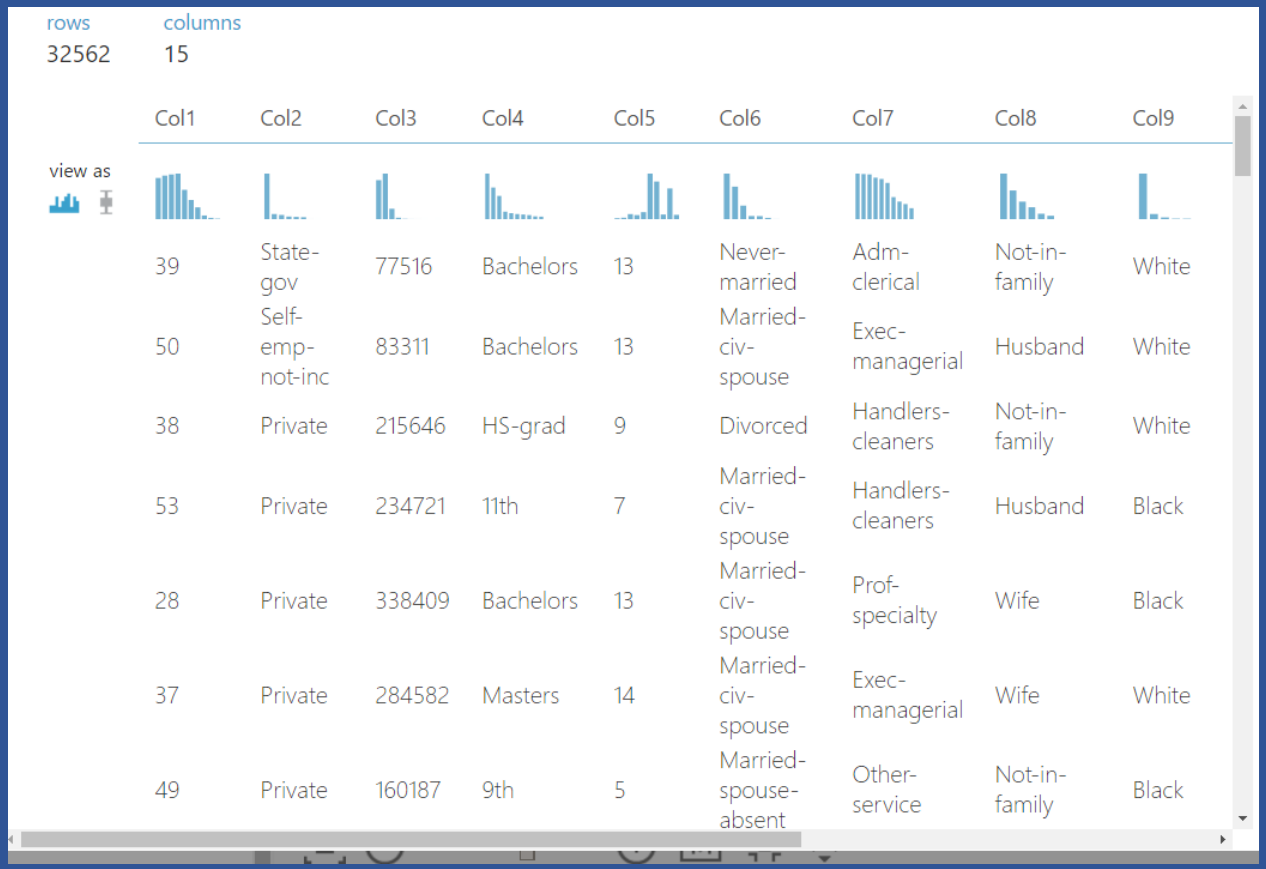
http://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data



R Script add column name

Import dataset from internet (Azure ML Studio)

Right-click result dataset (Dataset) at Import Data Visualize



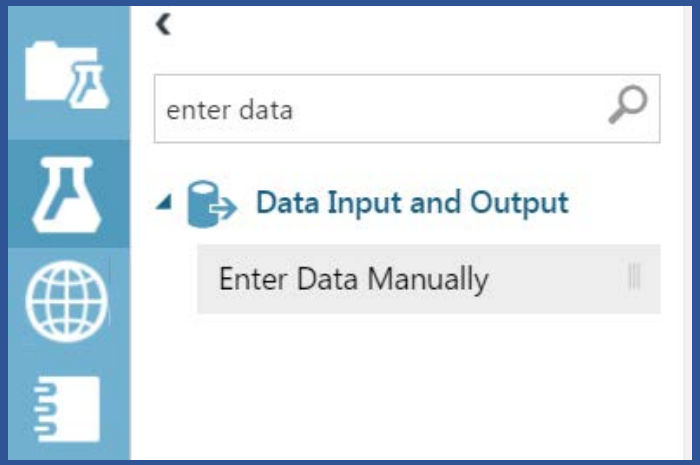
rows 32562 columns 15

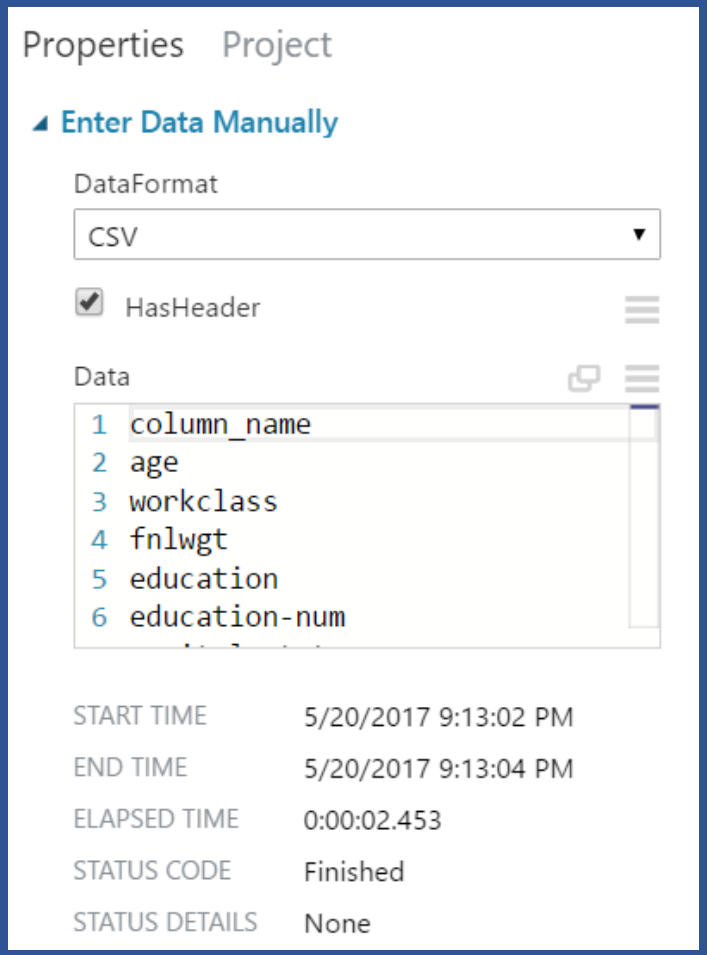
	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9
view as									
	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White
	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White
	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White
	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black
	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black
	37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White
	49	Private	160187	9th	5	Married-spouse-absent	Other-service	Not-in-family	Black

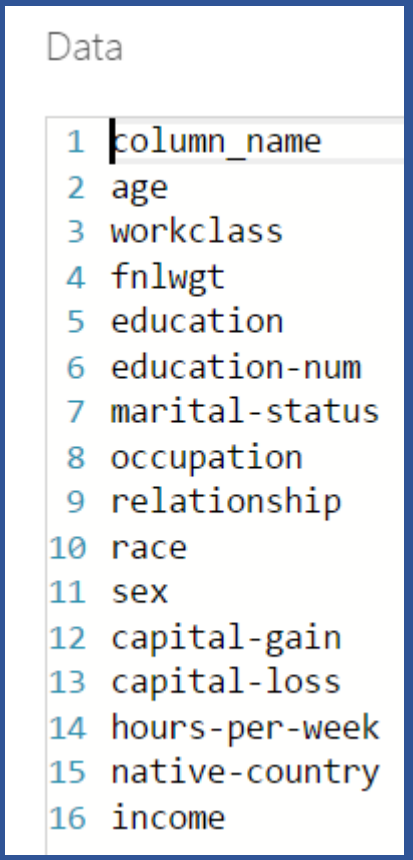
R Script add column name

Add column name Using R instructions (Azure ML Studio)

- Drag & drop Enter Data Manually from Data Input and Output to canvas
- Set properties



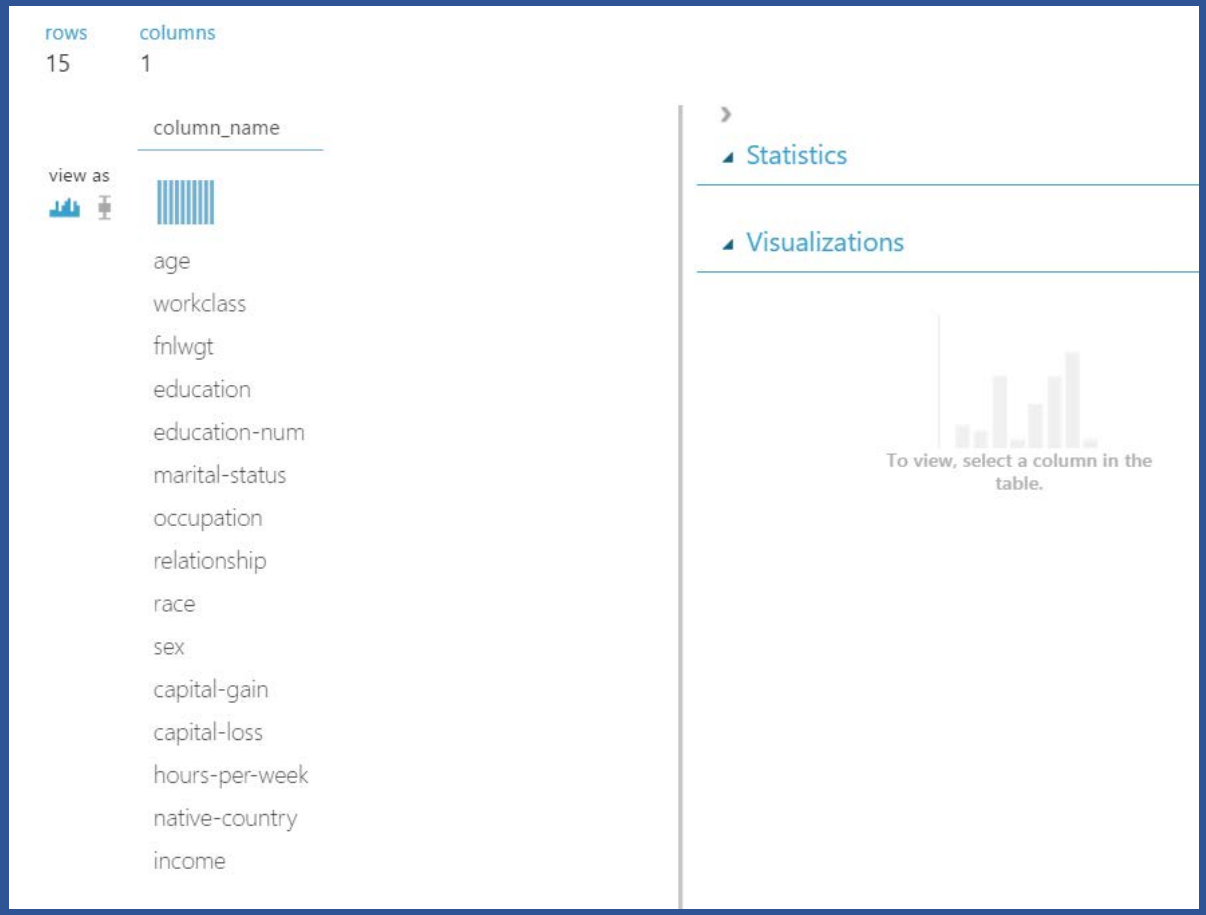




R Script add column name

Add column name Using R instructions (Azure ML Studio)

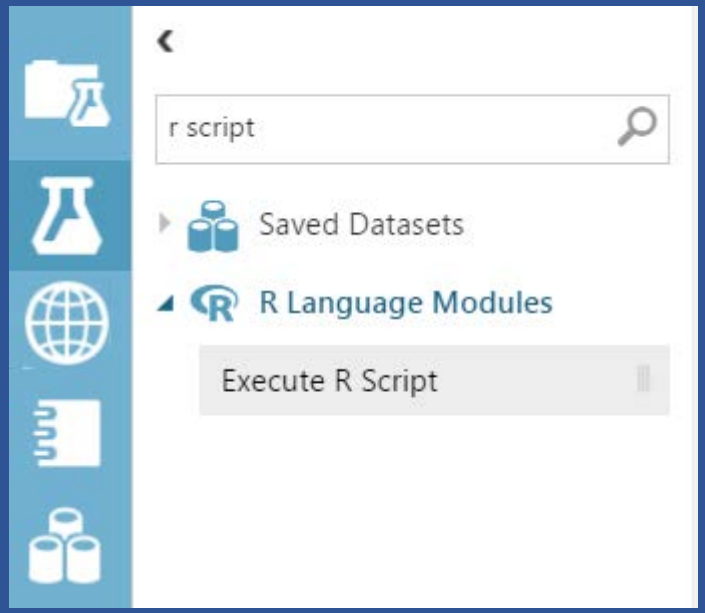
Right-click at Enter Data manually dataset (Dataset) Visualize

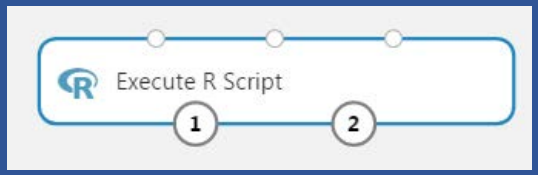


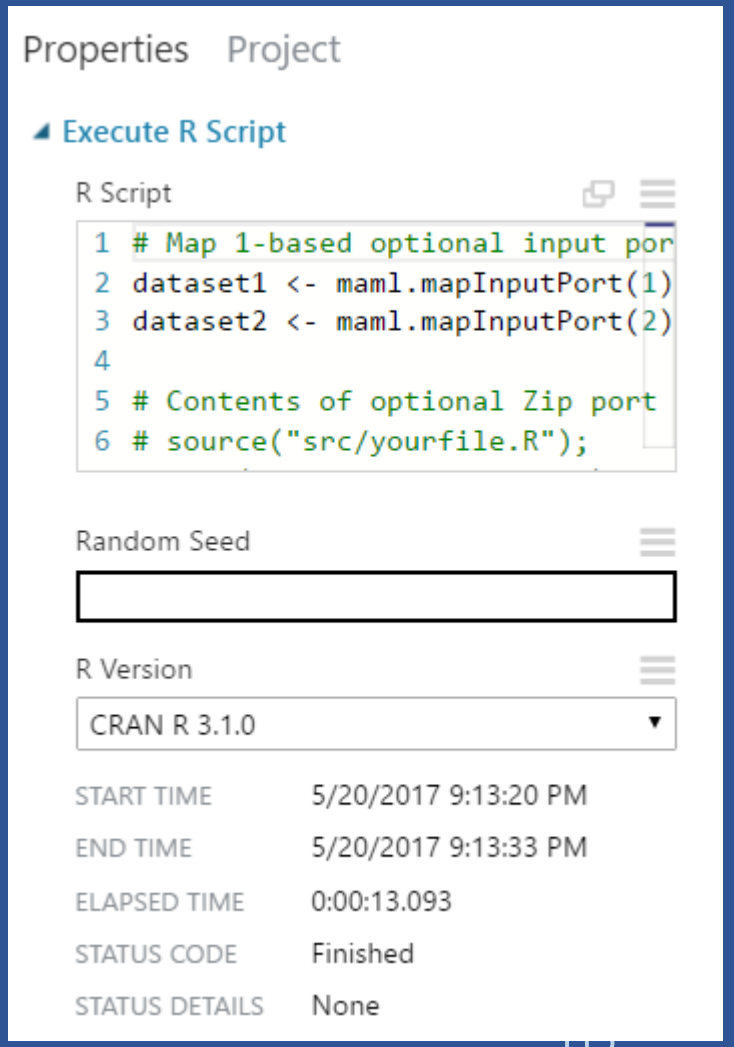
R Script add column name

Add column name Using R instructions (Azure ML Studio)

- Add Execute R Script module
- Configure R Script module
- Enter R Script







R Script add column name

Add column name Using R instructions (Azure ML Studio)

```
# Map 1-based optional input ports to variables
dataset1 <- maml.mapInputPort(1) # class: data.frame
dataset2 <- maml.mapInputPort(2) # class: data.frame

# Sample operation
colnames(dataset2) <- c(dataset1['column_name'])$column_name;
data.set = dataset2;

# Select data.frame to be sent to the output Dataset port
maml.mapOutputPort("data.set");
```

R Script add column name

Add column name Using R instructions (Azure ML Studio)

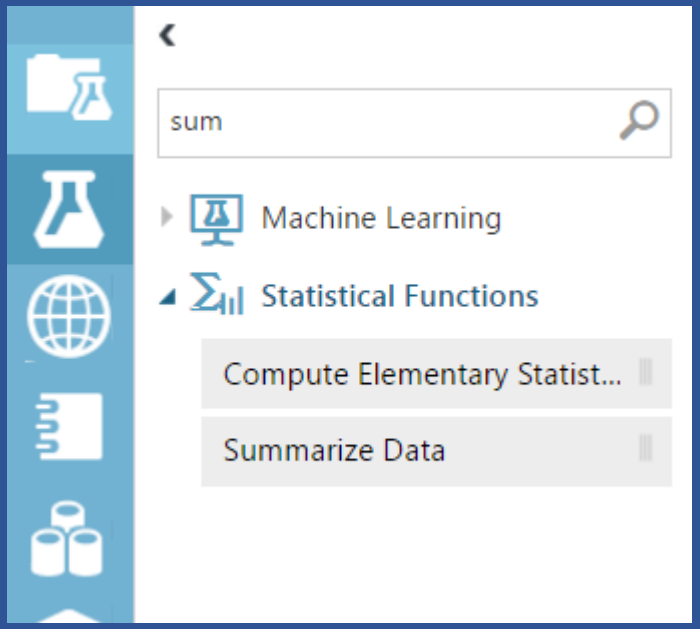
Visualize at the output of Execute R Script module

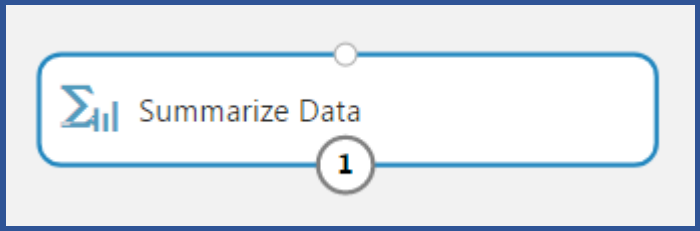
age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capital-gain	ca lo
39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0
50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0
38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0
53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0
28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0
37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White	Female	0	0

R Script add column name

Data Visualization (Azure ML Studio)

- Add Summarize data module
- Link to Execute R Script module





R Script add column name

Data Visualization (Azure ML Studio)

Summarize dataset visualization

Feature	Count	Unique Value Count	Missing Value Count	Min	Max	Mean	Mean Deviation	1st Quartile	Median	3rd Quartile
age	32561	73	1	17	90	38.581647	11.189182	28	37	48
workclass	30725	9	1837							
fnlwgt	32561	21648	1	12285	1484705	189778.366512	77608.21854	117827	178356	237051
education	32561	17	1							
education-num	32561	16	1	1	16	10.080679	1.903048	9	10	12
marital-status	32561	8	1							
occupation	30718	15	1844							
relationship	32561	7	1							
race	32561	6	1							
sex	32561	3	1							

R Script add column name

More Information

Import your training data into Azure Machine Learning Studio from various data sources

https://docs.microsoft.com/en-us/azure/machine-learning/machine-learning-data-science-import-data

This experiment

https://gallery.cortanaintelligence.com/Experiment/R-add-col-name