

# Introduction to Tableau Prep Builder

MRCB | s5 | 2019

Josep Curto | Professor, IE Business School | CEO, Delfos Research

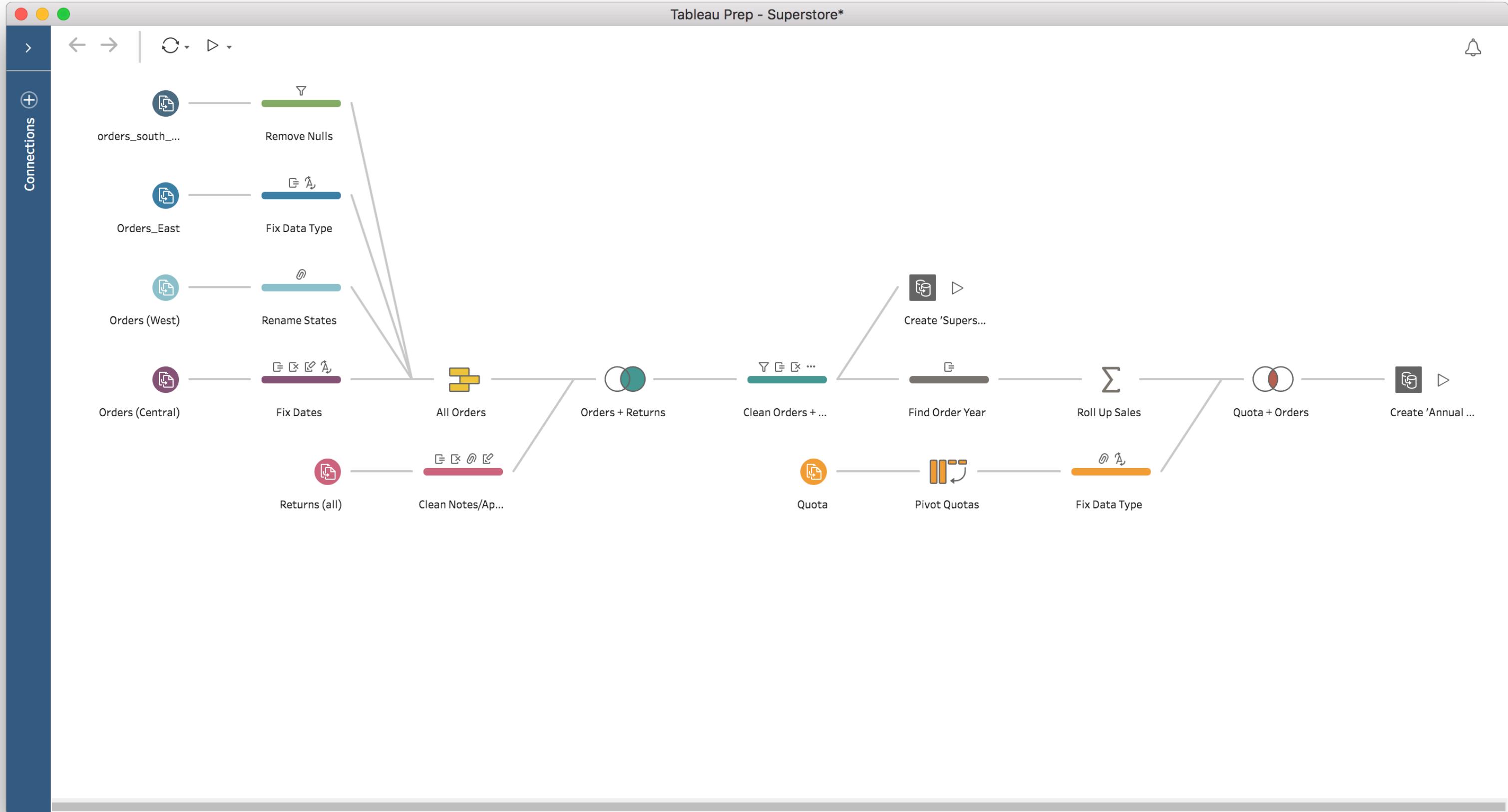
# Agenda

## Tableau Prep Builder

Introduction, Installation, Flows, Example

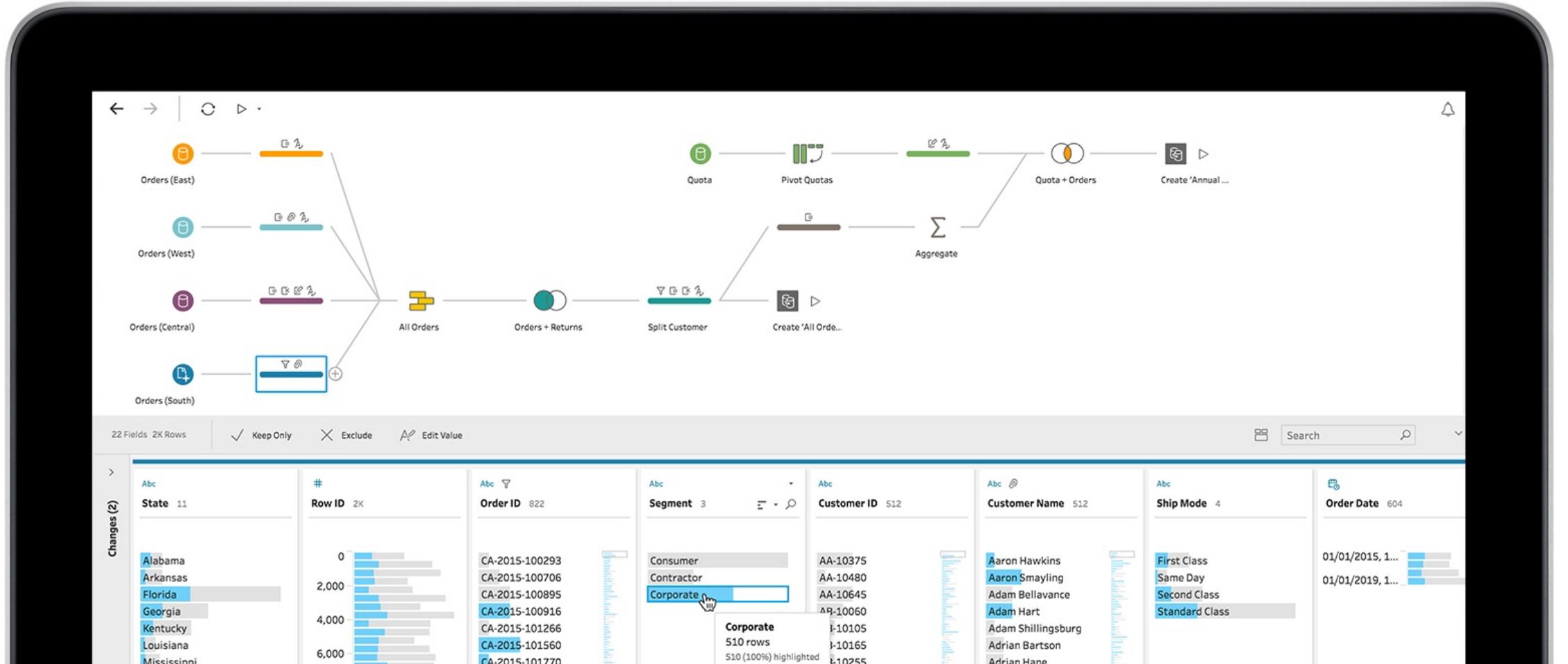
# Introduction

# What is Tableau Prep Builder



**Tableau Prep  
Builder is an **in-memory data processing tool** with multiple data connectors. It helps us to prepare the data and export the output as CSV or Tableau Data File.**

# Approach



- Focus on all-type of users (not only developers)
- Multiplatform
- Enterprise-ready, proprietary and based on licenses
- Based on data flows
- Easy to create, maintain and configure
- Well integrate in the Tableau Ecosystem

# Installation

# Can I install Tableau Prep Builder?

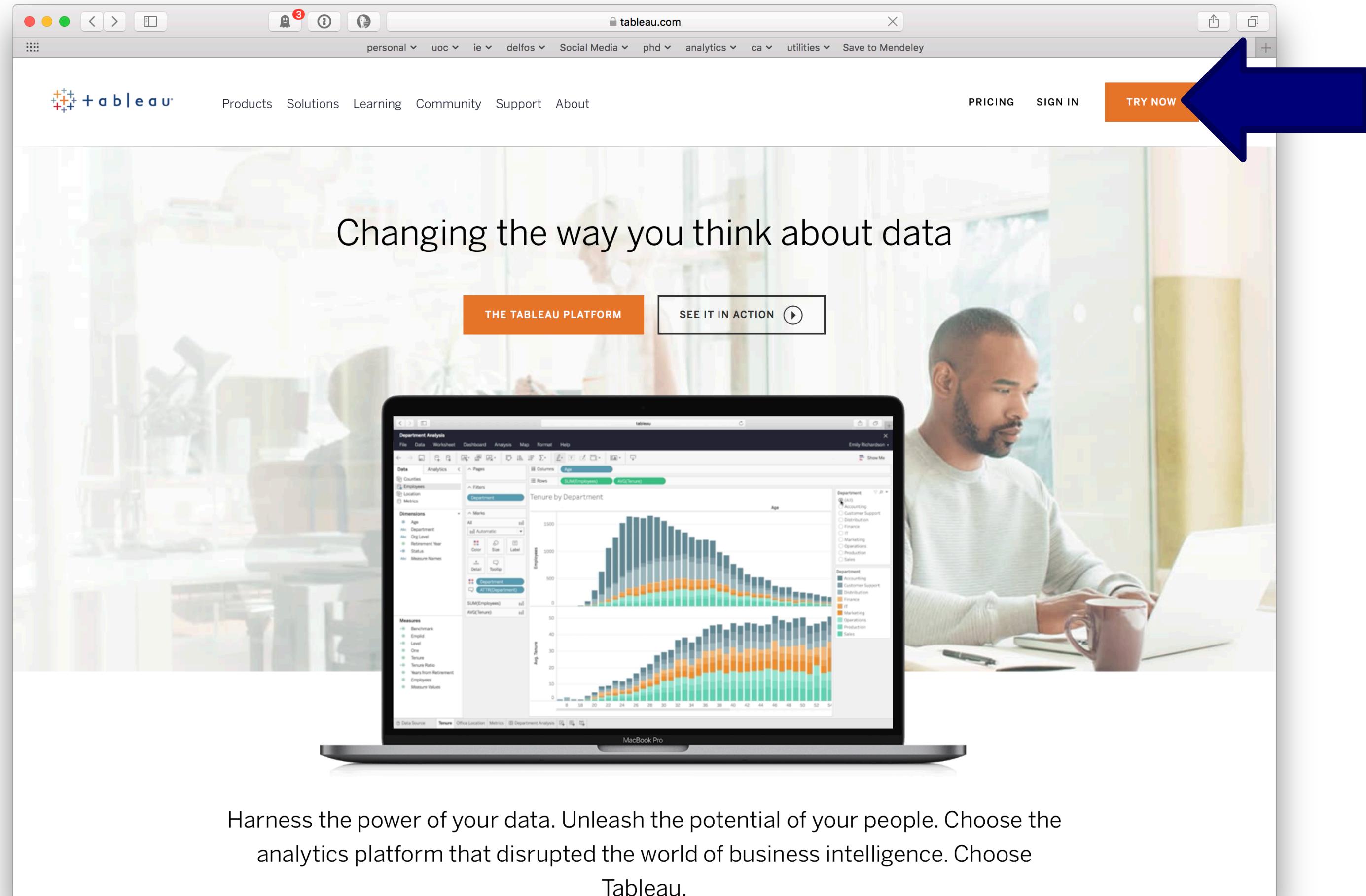
## Windows

- Windows 7 or newer (64 bit)
- Intel Core i3 or AMD Ryzen 3 Pro or faster
- 4 GB memory
- 2 GM minimum free disk space

## Mac

- Mac Os 10.11
- Intel Core i3 or faster
- 4 GB memory
- 2 GB minimum free disk space

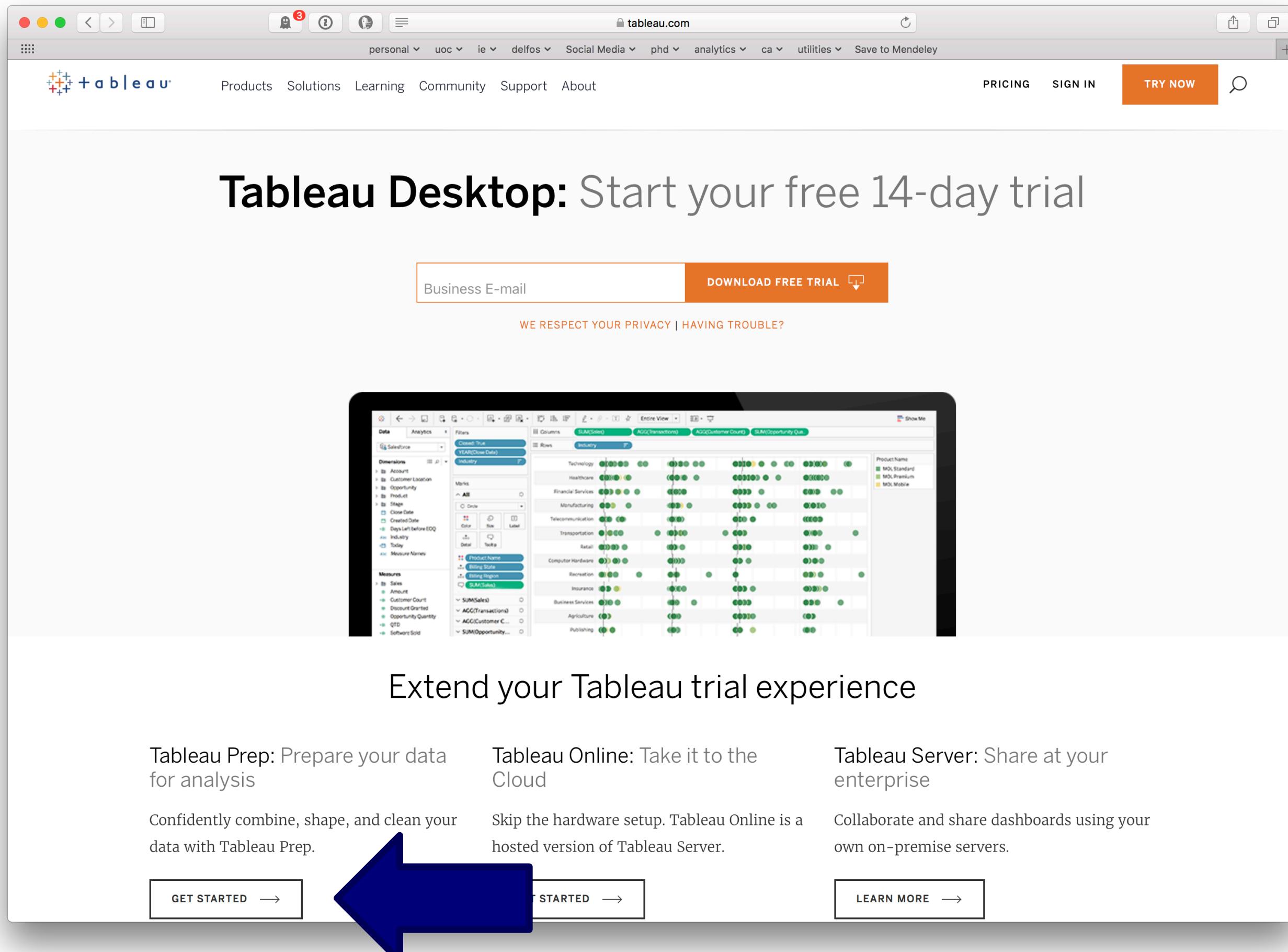
# How to install Tableau Prep Builder? (I)



- Go to: <https://www.tableau.com>
- Click **Try Me**

Note: At IE HST we use the provided landing page (check campus). Page may look different.

# How to install Tableau Prep Builder? (II)



- We are in: <https://www.tableau.com/products/trial>

- Click **GET Started**

*Note: Page may look different.*

# How to install Tableau Prep Builder? (III)

tableau.com

personal uoc ie delfos Social Media phd analytics ca utilities Save to Mendeley

Products Solutions Learning Community Support About PRICING SIGN IN TRY NOW

## Tableau Prep: Start your free trial

Business E-mail [START FREE TRIAL](#)

WE RESPECT YOUR PRIVACY | HAVING TROUBLE?

Changes (2) [VIEW SYSTEM REQUIREMENTS](#)

All of Tableau's products are Unicode-enabled and compatible with data stored in any language. The user interface and supporting documentation are in English,

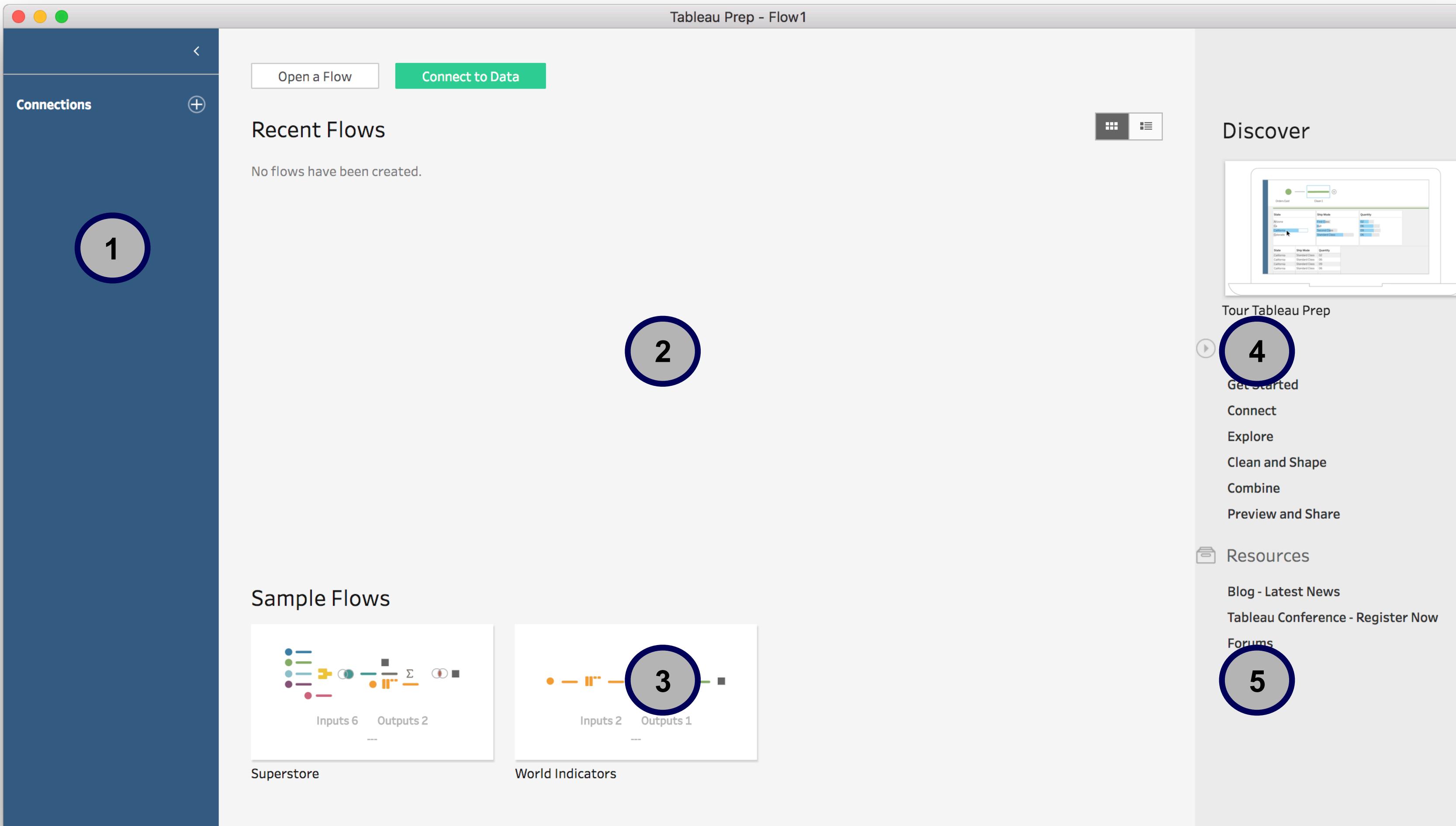
- Complete your email and **START FREE TRIAL**
- Download will start (speed will vary based on your bandwidth)
- Trail last 14 days
- Add your IE HST license

*Note: Page may look different.*

# (Working with) Flows

# Tableau Prep Builder Interface

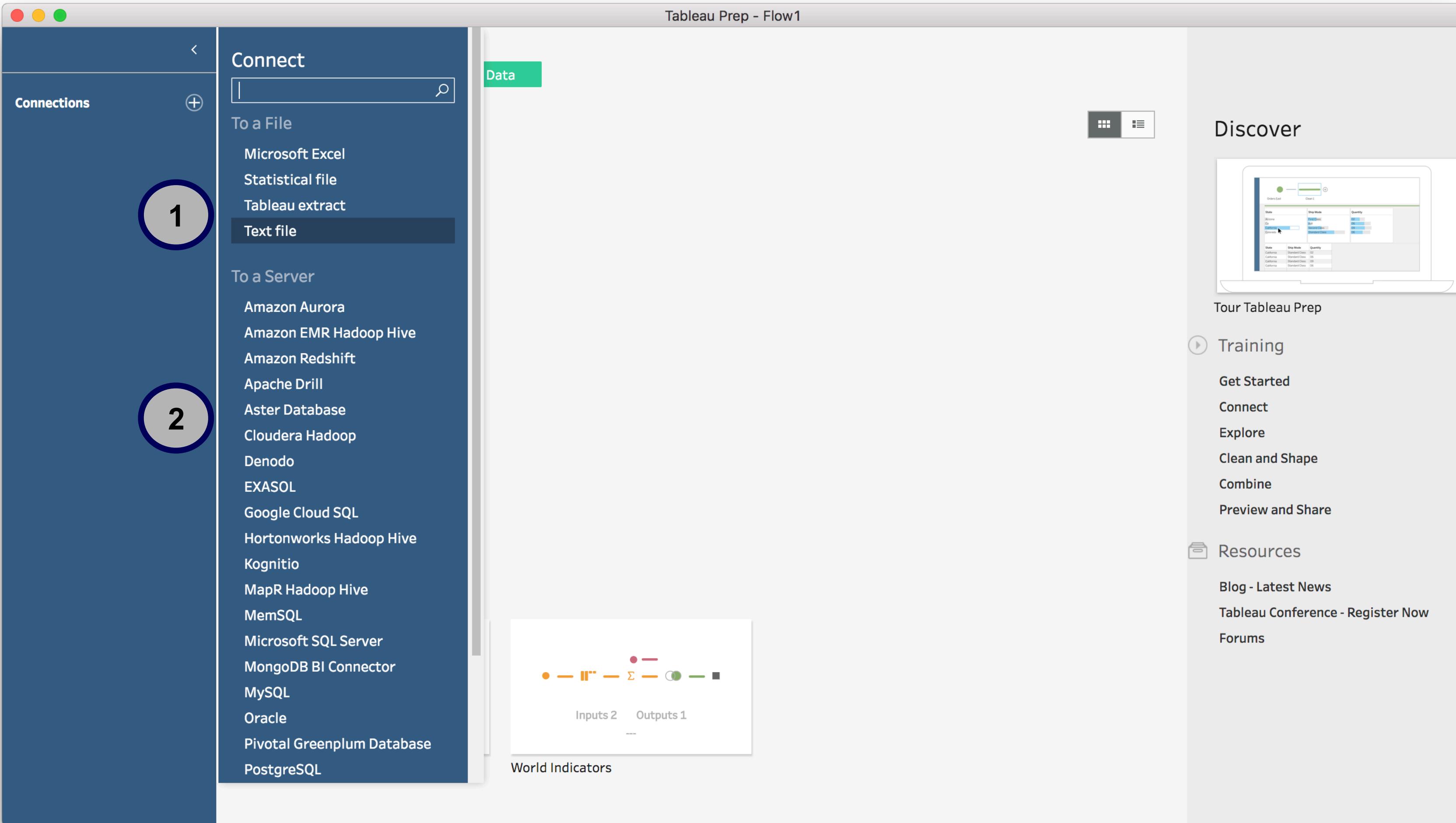
When we open the program, this is the first screen...



- 1 **Connections**
- 2 **Previous Flows**
- 3 **Sample Flows**
- 4 **Training**
- 5 **Resources & Updates**

# Data Connectors

There are multiple connectors:



1 **Files**

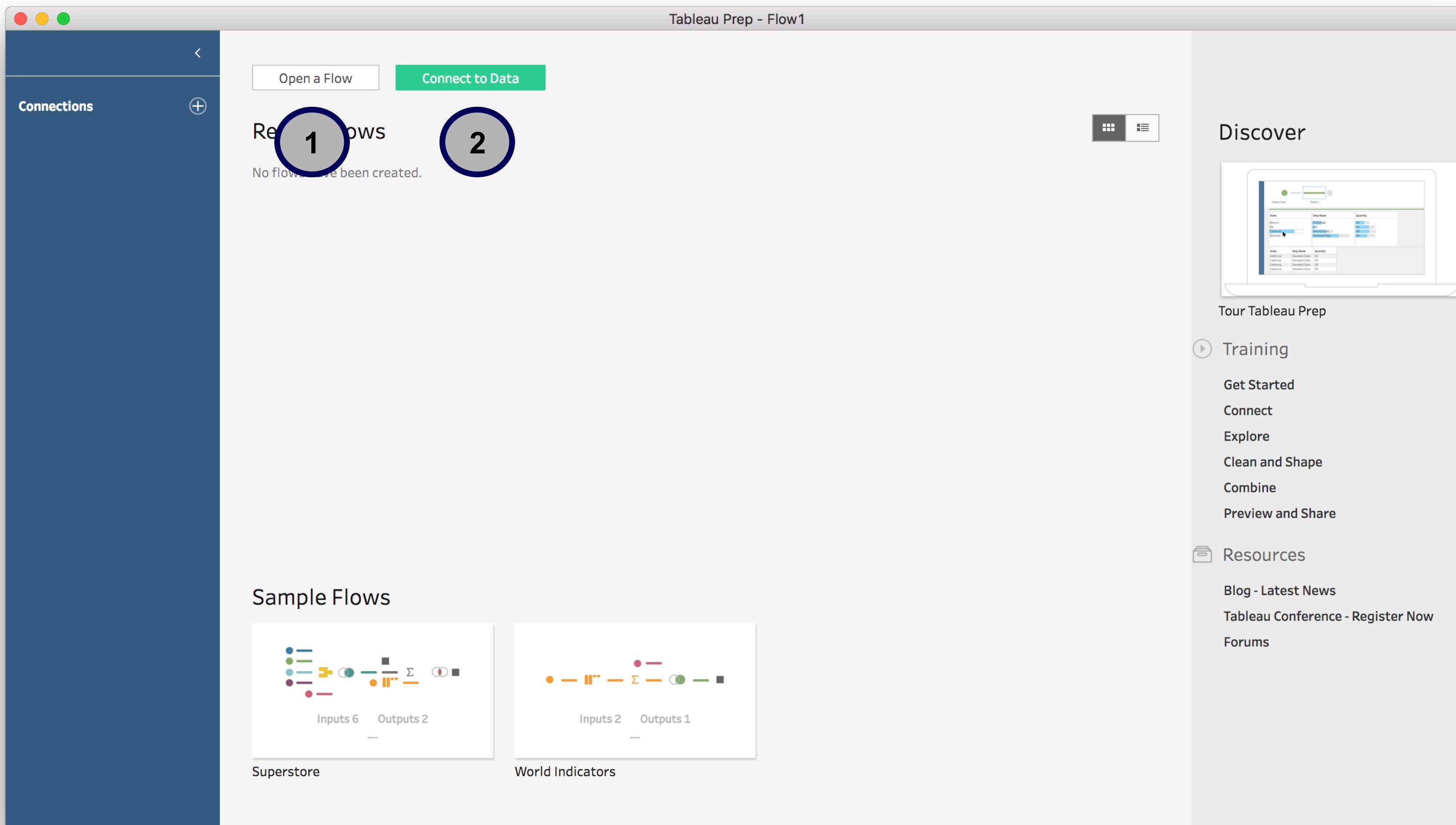
Including excel, text, tableau files and others

2 **Servers & Databases**

Including all type of databases, big data systems and enterprise and commercial applications

# How to work with Tableau Prep Builder

Two options:



## 1 Flows

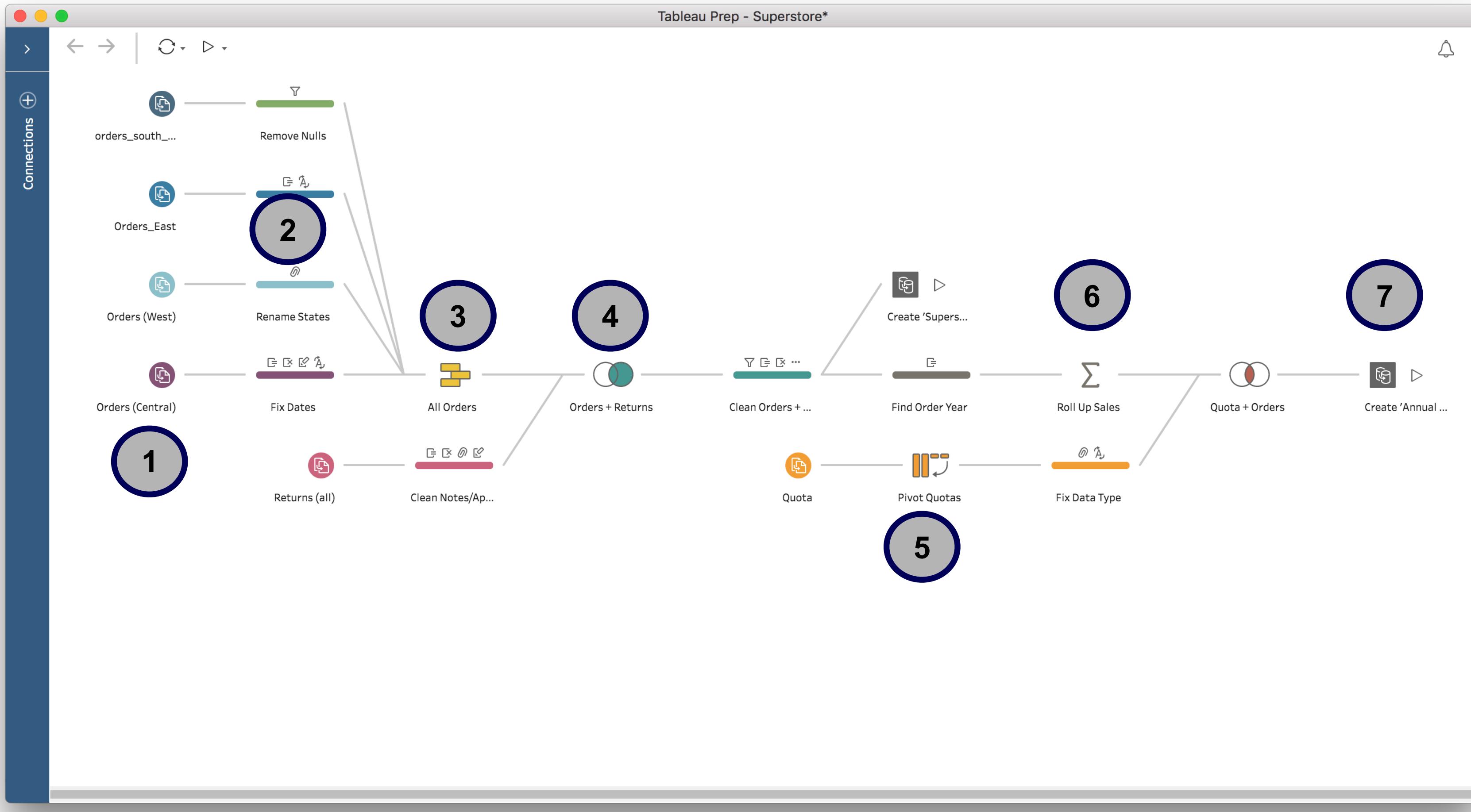
We open a previous flow (to execute or modify it)

## 2 Data

We open a data source (to create a new flow)

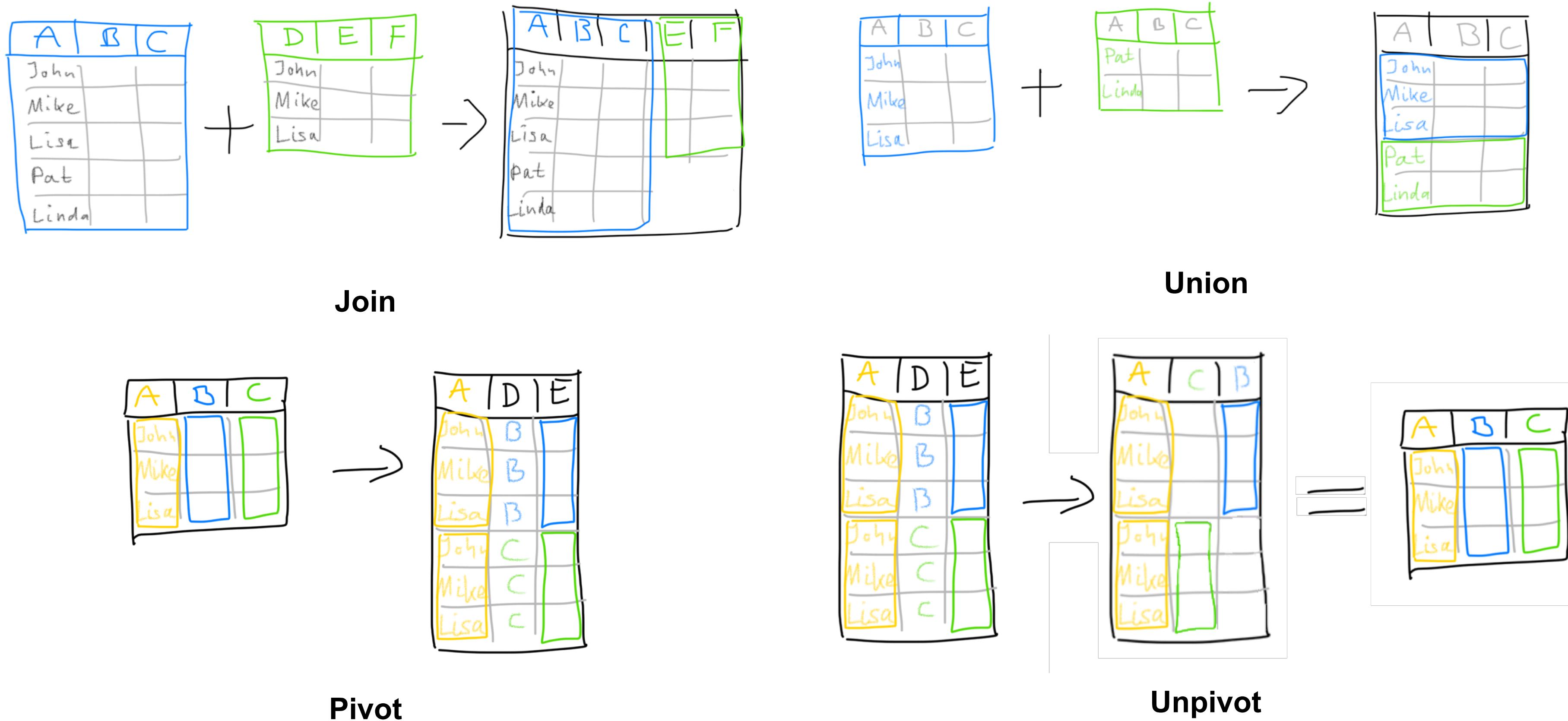
# What is a Flow?

A collection of steps that we apply to a data source. This is an ETL process (Extract, Transform and Load). We have several options:



- 1 Access data
- 2 Clean data
- 3 Union data (based on the same fields)
- 4 Join data (all, intersection,...)
- 5 Pivot data (create a pivot table)
- 6 Aggregate data (based on a metric or attribute)
- 7 Export

# Relevant steps in a flow

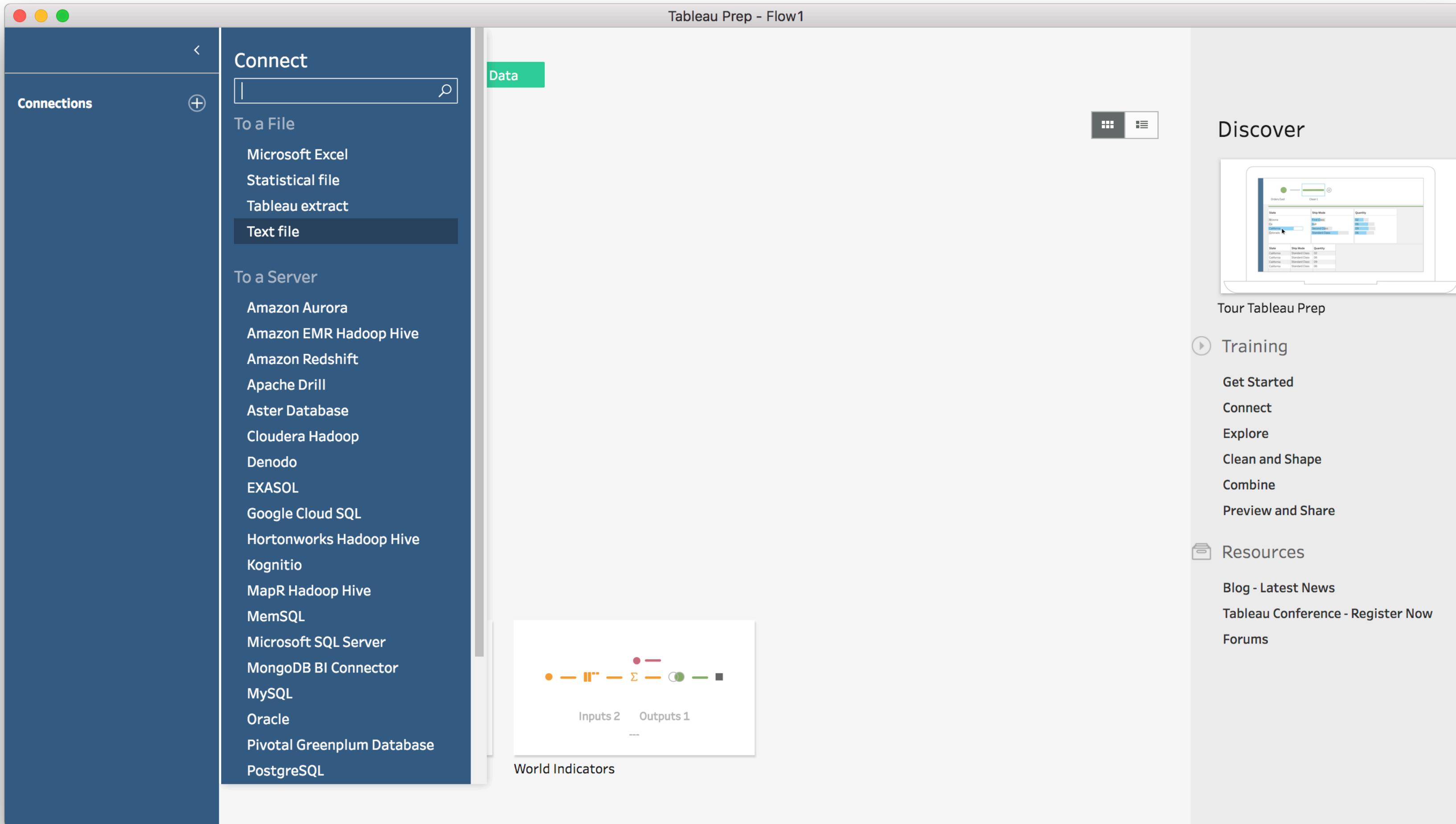


- When we have several data sources, the main options to combine them are: join, union, pivot and unpivot.
- The others steps are related to data cleaning and preparation.

# Example

(we recommend to check this link as well:  
[https://onlinehelp.tableau.com/current/prep/en-us/prep\\_welcome.htm](https://onlinehelp.tableau.com/current/prep/en-us/prep_welcome.htm))

# Acces data (I)



- We will prepare the dataset: **auto\_mpg.csv**
- Click on **Connections**
- Select **text file**

# Acces data (II)

The screenshot shows the Tableau Prep interface with the title 'Tableau Prep - Flow1\*'. On the left, the 'Connections' pane lists 'auto\_mpg.csv' as a 'Text file'. The 'Tables' pane shows 'auto\_mpg'. The main area is titled 'Input' and has three tabs: 'Text Settings' (selected), 'Multiple Files', and 'Data Sample'. Under 'Text Settings', the 'Connection' section shows 'auto\_mpg.csv' selected. The 'Text Options' section includes 'First line contains header' (radio button selected) and 'Generate field names automatically'. Below these are dropdowns for 'Field Separator' (set to 'Comma'), 'Text Qualifier' (set to 'Automatic'), 'Character Set' (set to 'UTF-8'), and 'Locale' (set to 'English (United States)'). To the right, the 'auto\_mpg' table is displayed with 9 fields selected. The table has columns: 'Type', 'Field Name', 'Original Field Name', 'Filters', and 'Sample Values'. The fields listed are: mpg, cylinders, displacement, horsepower, weight, acceleration, model\_year, origin, and car\_name.

Type	Field Name	Original Field Name	Filters	Sample Values
#	mpg	mpg		31.9, 44.3, 29
#	cylinders	cylinders		4
#	displacement	displacement		89, 90
#	horsepower	horsepower		71, 48, 70
#	weight	weight		1,925, 2,085, 1,937
#	acceleration	acceleration		14, 21.7, 14.2
#	model_year	model_year		79, 80, 76
#	origin	origin		2
Abc	car_name	car_name		vw rabbit custom, vw ra...

- For each connection we can change the options.
- In this case: *text settings, multiple files* y *data sample*.
- For *text settings*: format, separation, header, etc.

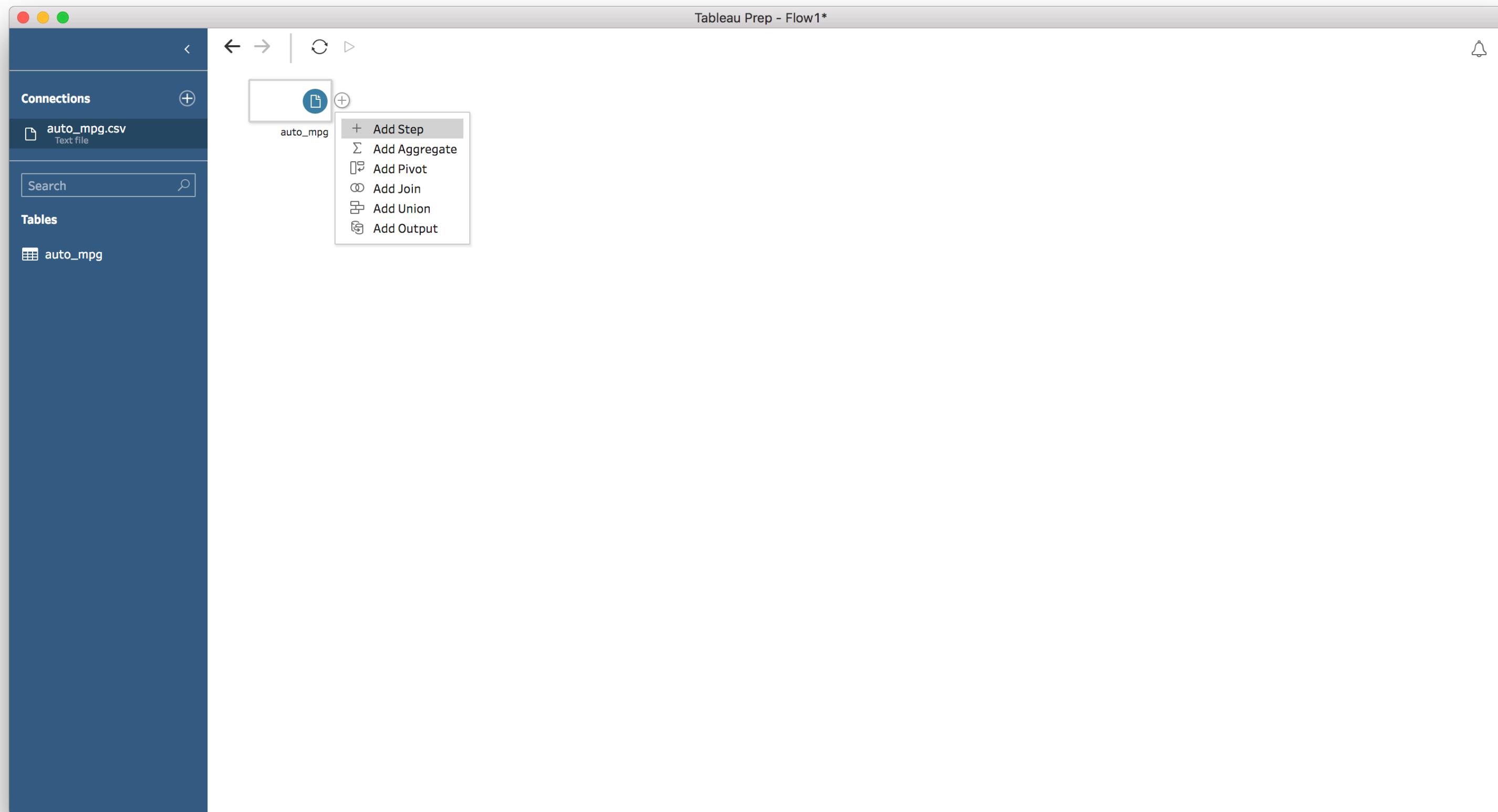
# Acces data (III)

The screenshot shows the Tableau Prep interface. On the left, the 'Connections' pane lists 'auto\_mpg.csv' as a Text file. The 'Tables' pane shows 'auto\_mpg'. The main area is titled 'Tableau Prep - Flow1\*' and contains a 'Data Sample' section for 'auto\_mpg'. It displays a table with 9 fields: mpg, cylinders, displacement, horsepower, weight, acceleration, model\_year, origin, and car\_name. The 'mpg' field has sample values 31.9, 44.3, 29. The 'cylinders' field has sample value 4. The 'displacement' field has sample value 89, 90. The 'horsepower' field has sample values 71, 48, 70. The 'weight' field has sample values 1,925, 2,085, 1,937. The 'acceleration' field has sample values 14, 21.7, 14.2. The 'model\_year' field has sample values 79, 80, 76. The 'origin' field has sample value 2. The 'car\_name' field has sample values vw rabbit custom, vw ra... The 'Input' section includes settings for 'Text Settings', 'Multiple Files', and 'Data Sample'. Under 'Data Sample', it says 'Fields selected: 9 of 9' and 'Select the fields to include in your flow. If you make changes to the data, the data source will be queried again.'

- If the file is small, we can use all the data (this case).
- If not use a sample (default or by number of rows)

# Data Preparation (I)

After loading the dataset, we have our step and we need to add the next one. In this step, we will start preparing the data. Click (+) and we will display a menu with all the options.

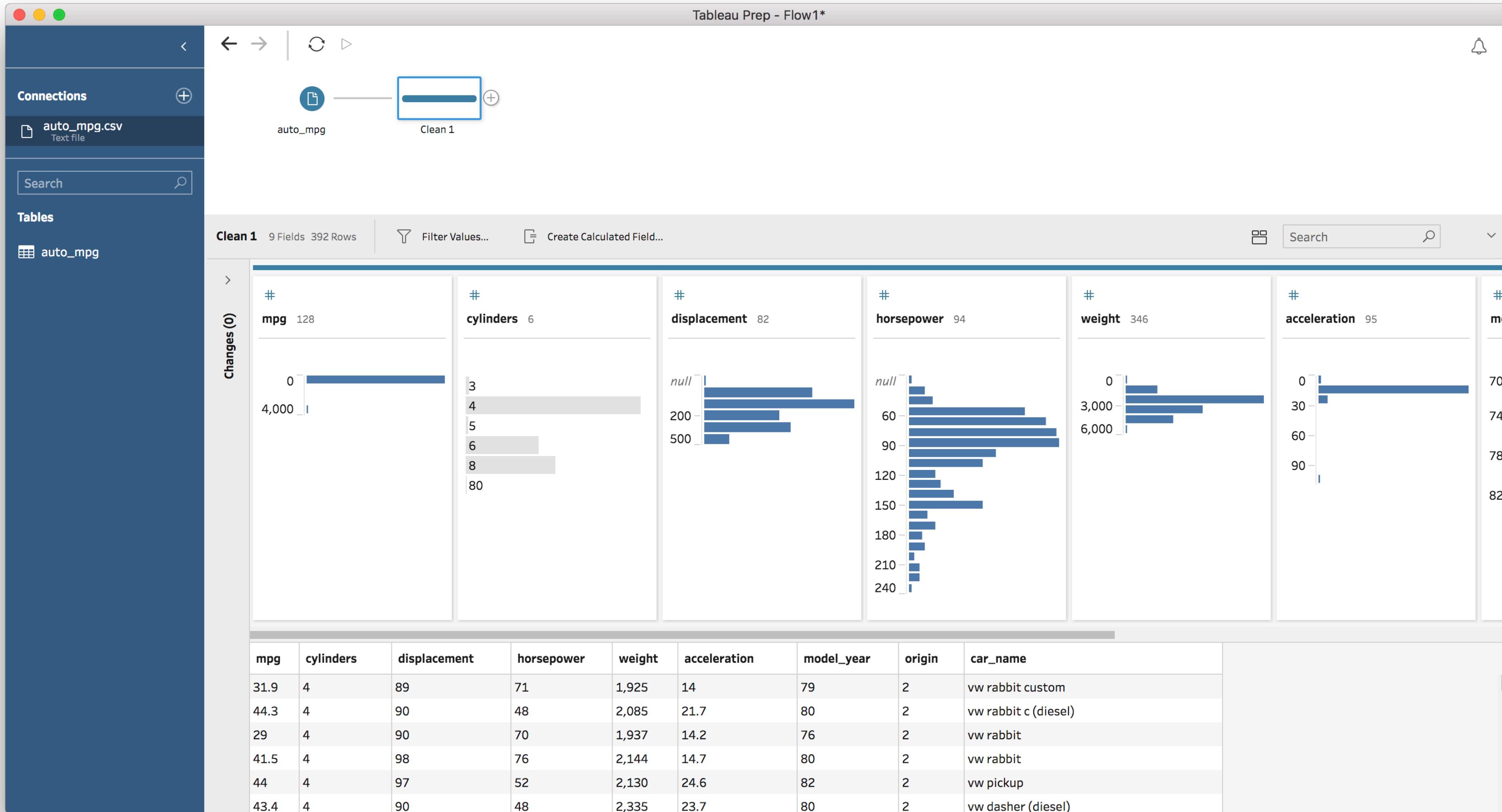


- We have six steps
- **Step:** to apply changes and cleansing functions to our data.
- **Aggregate, pivot, join y union:** to combine the flow of data
- **Output:** to export the flow to a new data set.

Check this link for more videos about all the steps in Tableau Prep Builder: <https://www.tableau.com/learn/training#prep>

# Data Preparation (II)

As starting point, we will prepare the data in auto\_mpg.csv



- When selecting the first *step*, we will have the distribution per field. This can help us to spot error to solve and transformations to apply.
- One thing that we can do is create calculated fields.

# Data Preparation (III)

Using *calculated field*, we can create new fields in our data set.

The screenshot shows the Tableau Prep interface with a flow from 'auto\_mpg' to 'Clean 1' and then to 'Output'. The 'Clean 1' step is selected. In the 'Edit Field' dialog for 'Region', the formula is defined as:

```
IF [origin]=1  
THEN 'US'  
ELSEIF [origin]=2  
THEN 'Europe'  
ELSE 'Japan'  
END
```

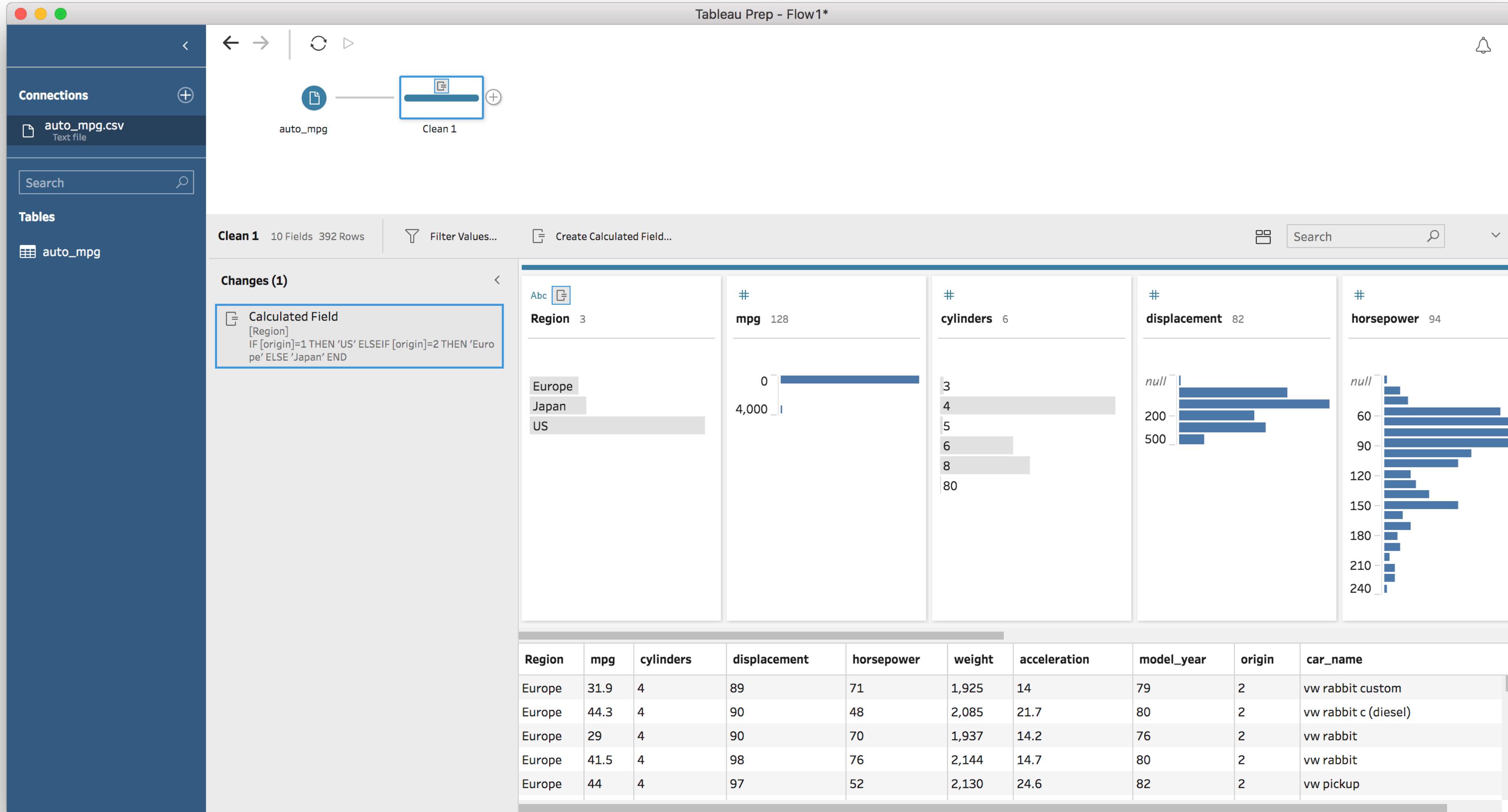
The 'Reference' sidebar shows the definition of the **ABS(number)** function. Below the formula, a preview of the data shows columns for **displacement** and **horsepower**, and a small table of car data with columns **origin**, **model\_year**, **car\_name**.

- Let's create a new field called **region** based on **origin**.

**IF [origin]=1  
THEN 'US'  
ELSEIF [origin]=2  
THEN 'Europe'  
ELSE 'Japan'  
END**

# Data Preparation (IV)

Every time that will apply a transformation, we will find it in *changes* as a sequence:



# Data Preparation (V)

The screenshot shows the Tableau Prep interface. On the left, the 'Connections' pane lists 'auto\_mpg.csv' as a Text file. The 'Tables' pane shows 'auto\_mpg'. In the center, a flow diagram has a source node 'auto\_mpg' connected to a transformation node. A tooltip for the transformation node says 'Clean 1 10 Fields 392 Rows'. Below it, the 'Changes (1)' section shows a 'Calculated Field' named 'Year' with the formula `STR([model_year] + 1900)`. A detailed tooltip for 'ABS(number)' is open, explaining it returns the absolute value of a number. At the bottom, a preview of the data shows columns: Region, mpg, cylinders, displacement, horsepower, weight, acceleration, model\_year, origin, and car\_name. The 'model\_year' column contains values like 70, 76, 80, etc.

- We can create "**Year**" from "Model Year" using **calculated field**:

**STR([Model Year] + 1900)**

# Data Preparation (VI)

The screenshot shows the Tableau Prep interface with the following details:

- Connections:** auto\_mpg.csv (Text file) is connected to Clean 1.
- Tables:** auto\_mpg
- Changes (4):**
  - Calculated Field [Region]: IF [origin]=1 THEN 'US' ELSEIF [origin]=2 THEN 'Europe' ELSE 'Japan' END
  - Calculated Field [Year]: STR([model\_year]+1900)
  - Calculated Field [car\_name - Split 1]: TRIM(SPLIT([car\_name], " ", 1))
  - Calculated Field [car\_name - Split 2]: INT(SPLIT([car\_name], " ", 2))
- Visualizations:** Four histograms are displayed:
  - car\_name - Split 1: Shows car brands like amc, audi, bmw, buick, cadillac, capri, chevrolet, chevy, chrysler, datsun, and dodge.
  - car\_name - Split 2: Shows years from 1970 to 1981.
  - Year: Shows the count of 13 entries.
  - Region: Shows the count of 3 entries (Europe, Japan, US).
  - mpg: Shows the count of 128 entries ranging from 0 to 4,000.
- Table View:** A preview table shows the first few rows of the cleaned data, including columns: car\_name - Split 1, car\_name - Split 2, Year, Region, mpg, cylinders, displacement, horsepower, weight, and acceleration.

- We can create “Brand” applying *split* to “Car Name”.
- Two columns are generated: split 1 and 2.
- We need the first one only.
- We will delete the second one with *Remove Field*

# Data Preparation (VII)

The screenshot shows the Tableau Prep interface with a flow named 'Flow1'. The flow starts with a connection to 'auto\_mpg.csv' and leads to a step named 'Clean 1'. The 'Changes' pane on the left lists seven modifications made to the 'Clean 1' step:

- Calculated Field [Region] IF [origin]=1 THEN 'US' ELSEIF [origin]=2 THEN 'Europe' ELSE 'Japan' END
- Calculated Field [Year] STR([model\_year]+1900)
- Calculated Field [car\_name - Split 1] TRIM(SPLIT([car\_name], " ", 1))
- Calculated Field [car\_name - Split 2] INT(SPLIT([car\_name], " ", 2))
- Rename Field [brand] From [car\_name - Split 1] to [brand]
- Make Uppercase [Year] All values changed to uppercase
- Rename Field [year] From [Year] to [year]

The main workspace displays three data sources: 'brand' (37 rows), 'car\_name - Split 2' (24 rows), and 'year' (13 rows). A context menu is open over the 'Region' field in the 'Clean 1' step, showing options like 'Filter', 'Group and Replace', 'Clean', 'Split Values', 'View State', 'Rename Field', 'Create Calculated Field...', and 'Remove Field'.

- If we make a mistake with a field name we can change it (using *rename field*).
- In every menu we can find several options that can help us.

# Data Preparation (VIII)

The screenshot shows the Tableau Prep interface with a flow named 'Clean 1'. The flow consists of a connection to 'auto\_mpg.csv' and a step labeled 'Clean 1'. The 'Changes (9)' pane on the left lists various data transformations:

- Calculated Field [Region] IF [origin]=1 THEN 'US' ELSEIF [origin]=2 THEN 'Europe' ELSE 'Japan' END
- Calculated Field [Year] STR([model\_year]+1900)
- Calculated Field [car\_name - Split 1] TRIM(SPLIT([car\_name], " ", 1))
- Calculated Field [car\_name - Split 2] INT(SPLIT([car\_name], " ", 2))
- Rename Field [brand] From [car\_name - Split 1] to [brand]
- Make Uppercase [year] All values changed to uppercase
- Rename Field [year] From [Year] to [year]
- Rename Field [region] From [Region] to [region]
- Remove Field [car\_name - Split 2]

The main workspace displays four columns of data:

- brand**: 37 unique values including amc, audi, bmw, buick, cadillac, capri, chevrolet, chevy, chrysler, datsun, dodge.
- region**: 3 unique values including Europe, Japan, US.
- mpg**: 128 unique values ranging from 0 to 4,000.
- cylinders**: 6 unique values including 3, 4, 5, 6, 8, 80.

A tooltip for the 'brand' column indicates a range from 1980 to 1981.

mpg	cylinders	displacement	horsepower	weight	acceleration	model_year	origin	car_name
31.9	4	89	71	1,925	14	79	2	vw rabbit custom
44.3	4	90	48	2,085	21.7	80	2	vw rabbit c (diesel)
29	4	90	70	1,937	14.2	76	2	vw rabbit
41.5	4	98	76	2,144	14.7	80	2	vw rabbit
44	4	97	52	2,130	24.6	82	2	vw pickup

- After deleting split 2 borrar split 2, we have detected that there are some errors in the brand names. We must correct them.
- We will create another calculated field.

# Data Preparation (IX)

The screenshot shows the Tableau Prep interface with a flow named 'Flow1'. The flow starts with a connection to 'auto\_mpg.csv' and a step labeled 'Clean 1'. A modal window titled 'Add Field' is open, showing the creation of a calculated field named 'car\_brand'. The calculation is defined as:

```

if ([Brand]='toyouta') THEN
'toyota'
ELSEIF ([Brand]='mercedes') THEN
'mercedes benz'
ELSEIF ([Brand]='mercedes-benz') THEN
'mercedes benz'
ELSEIF ([Brand]='maxda') THEN
'mazda'
ELSEIF ([Brand]='chevy') THEN
'chevrolet'
ELSEIF ([Brand]='chevroelt') THEN
'chevrolet'
ELSEIF ([Brand]='vw') THEN
'volkswagen'
ELSEIF ([Brand]='vokswagen') THEN
'volkswagen'
ELSEIF ([Brand]='capri') THEN
'mercury'
ELSE [Brand]
END

```

The 'Reference' dropdown is set to 'All', and the 'ABS(number)' function is selected. A tooltip for 'ABS' states: 'Returns the absolute value of the given number.' Below the modal, a preview of the data shows columns like 'origin' and 'car\_name' with values such as 'vw rabbit custom' and 'vw pickup'.

- We will create "Card Brand" using *calculated field*:

```

if ([Brand]='toyouta') THEN
'toyota'
ELSEIF ([Brand]='mercedes') THEN
'mercedes benz'
ELSEIF ([Brand]='mercedes-benz') THEN
'mercedes benz'
ELSEIF ([Brand]='maxda') THEN
'mazda'
ELSEIF ([Brand]='chevy') THEN
'chevrolet'
ELSEIF ([Brand]='chevroelt') THEN
'chevrolet'
ELSEIF ([Brand]='vw') THEN
'volkswagen'
ELSEIF ([Brand]='vokswagen') THEN
'volkswagen'
ELSEIF ([Brand]='capri') THEN
'mercury'
ELSE [Brand]
END

```

# Data Preparation (X)

The screenshot shows the Tableau Prep interface with a flow named 'Flow1'. On the left, the 'Connections' pane shows a connection to 'auto\_mpg.csv'. The main workspace displays a flow with a source node 'auto\_mpg' connected to a 'Clean 1' node. The 'Changes (10)' pane on the left lists various transformations applied to the data:

- [Region] IF [origin]=1 THEN 'US' ELSEIF [origin]=2 THEN 'Europe' ELSE 'Japan' END
- Calculated Field [Year] STR([model\_year]+1900)
- Calculated Field [car\_name-Split1] TRIM(SPLIT([car\_name], " ", 1))
- Calculated Field [car\_name-Split2] INT(SPLIT([car\_name], " ", 2))
- Rename Field [brand] From [car\_name-Split1] to [brand]
- Make Uppercase [Year] All values changed to uppercase
- Rename Field [year] From [Year] to [year]
- Rename Field [region] From [Region] to [region]
- Remove Field [car\_name-Split2]
- Calculated Field [car\_brand] if ([brand]='toyota') THEN 'toyota' ELSEIF ([brand]='mercedes') THEN 'mercedes benz' ELSEIF

The workspace contains four preview panes: 'car\_brand' (29 rows), 'brand' (37 rows), 'year' (13 rows), and 'region' (3 rows). Below these is a preview of the final 'Clean 1' stage, which includes columns: car\_brand, brand, year, region, mpg, cylinders, displacement, horsepower, weight, acceleration, model\_year, and origin. The 'brand' column is highlighted with a blue border.

- After creating **car\_brand**, we can delete **brand** (as you can imagine only if we don't need this field anymore).

# Data Output (I)

The screenshot shows the Tableau Prep interface with a flow named 'Flow1'. On the left, the 'Connections' pane shows a single connection to 'auto\_mpg.csv'. The main workspace displays a 'Clean 1' step, which is connected to the input 'auto\_mpg'. A context menu is open over the 'Clean 1' step, showing options like 'Add Step', 'Add Aggregate', etc. Below the step, there are five preview panes for fields: 'car\_brand' (29 unique values), 'year' (13 unique values), 'region' (3 unique values), 'mpg' (128 unique values), and 'cylinders' (6 unique values). The 'Changes (15)' pane on the left lists the modifications made during the clean step, including Rename Field, Make Uppercase, and Calculated Field operations. At the bottom, a preview of the cleaned data is shown in a table with columns: mpg, cylinders, displacement, horsepower, weight, acceleration, model\_year, origin, and car\_name.

- After the data preparation, we must create the output file and flow.
- That means a new file that contains all the changes (and we will use during our analysis).

# Data Output (II)

The screenshot shows the Tableau Prep interface. On the left, there's a sidebar with 'Connections' containing 'auto\_mpg.csv' (Text file) and a 'Tables' section with 'auto\_mpg'. The main workspace shows a flow starting from 'auto\_mpg', going through a 'Clean 1' step, and ending at 'Output'. A 'Save As' dialog box is open over the workspace, prompting the user to save the output as 'auto\_mpg.hyper'. The dialog also includes fields for 'Tags' and 'Where' (set to 'Datasources'). Below the dialog, the 'Output' pane displays a preview of the data with 12 fields: car\_brand, year, region, mpg, cylinders, displacement, horsepower, weight, acceleration, model\_year, origin, and car\_name. The data preview shows rows for various cars from 1972 to 1982, primarily from Europe, with values for mpg ranging from 17 to 43.1.

- We must select the output format (recommended hyper).
- Remember to choose a proper name and a folder.

# Data Output (III)

The screenshot shows the Tableau Prep interface. On the left, the 'Connections' pane lists a single connection to 'auto\_mpg.csv'. The 'Tables' pane shows the 'auto\_mpg' table. The main workspace displays a flow starting with a 'Clean 1' step connected to an 'Output' step. A modal dialog titled 'Tableau Prep - Flow1\*' is open, prompting to 'Save As: auto\_mpg.tfl'. The 'Format:' dropdown is set to 'Tableau Flow Files'. Below the flow, a preview window titled 'Save to auto\_mpg.hyper' shows a sample of the data from the 'auto\_mpg' table.

car_brand	year	region	mpg	cylinders	displacement	horsepower	weight	acceleration	model_year	origin	car_name
volkswagen	1979	Europe	31.9	4	89	71	1,925	14	79	2	vw rabbit
volkswagen	1980	Europe	44.3	4	90	48	2,085	21.7	80	2	vw rabbit
volkswagen	1976	Europe	29	4	90	70	1,937	14.2	76	2	vw rabbit
volkswagen	1980	Europe	41.5	4	98	76	2,144	14.7	80	2	vw rabbit
volkswagen	1982	Europe	44	4	97	52	2,130	24.6	82	2	vw pickup
volkswagen	1980	Europe	43.4	4	90	48	2,335	23.7	80	2	vw dashe
volvo	1981	Europe	30.7	6	145	76	3,160	19.6	81	2	volvo dies
volvo	1978	Europe	17	6	163	125	3,140	13.6	78	2	volvo 264
volvo	1976	Europe	20	4	130	102	3,150	15.7	76	2	volvo 245
volvo	1975	Europe	22	4	121	98	2,945	14.5	75	2	volvo 244
volvo	1972	Europe	18	4	121	112	2,933	14.5	72	2	volvo 145
volvo	1973	Europe	19	4	121	112	2,868	15.5	73	2	volvo 144
volkswagen	1972	Europe	23	4	97	54	2,254	23.5	72	2	volkswag
volkswagen	1973	Europe	26	4	97	46	1,950	21	73	2	volkswag
volkswagen	1978	Europe	31.5	4	89	71	1,990	14.9	78	2	volkswag
volkswagen	1982	Europe	36	4	105	74	1,980	15.3	82	2	volkswag
volkswagen	1978	Europe	43.1	4	90	48	1,985	21.5	78	2	volkswag
volkswagen	1977	Europe	29	4	97	78	1,940	14.5	77	2	volkswag

- We must save the flow as well.
- This way we can modify the flow or share it with the rest of the team.



