

Introduction to Tableau Prep Builder

Josep Curto | Professor, IE Business School

Rev. 2020 v1

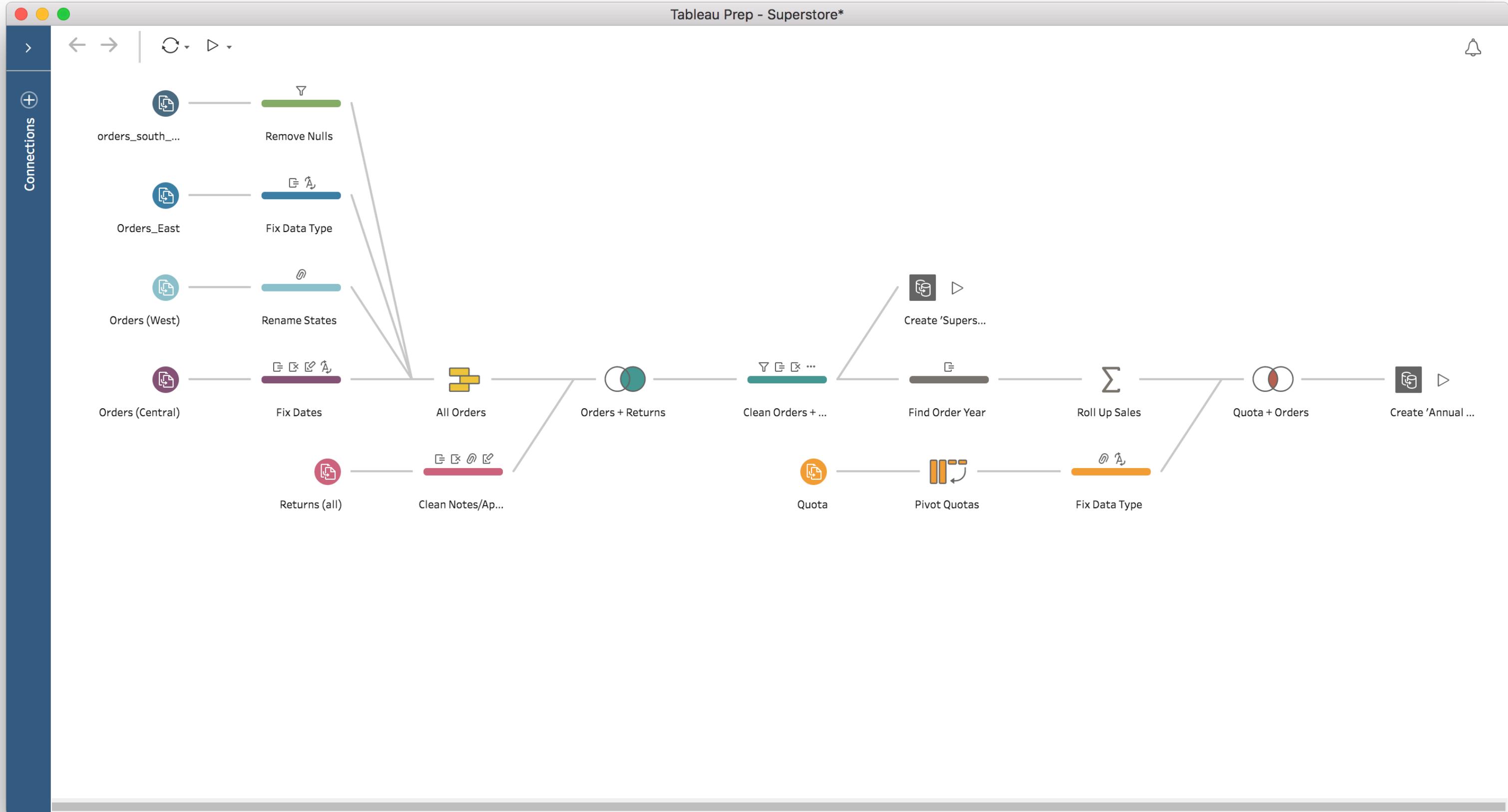
Agenda

Tableau Prep Builder

Introduction, Installation, Flows, Example

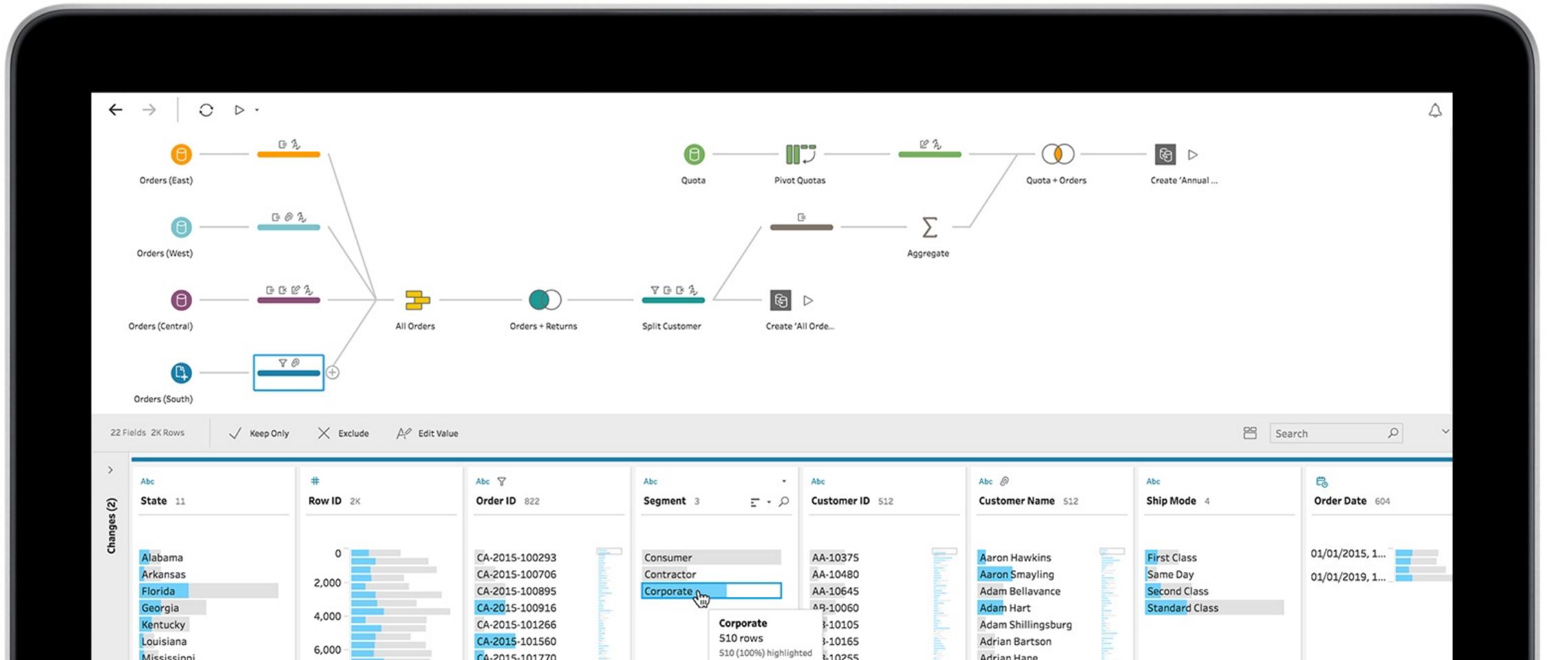
Introduction

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

Generic Installation

Can I install Tableau Prep Builder?

Windows

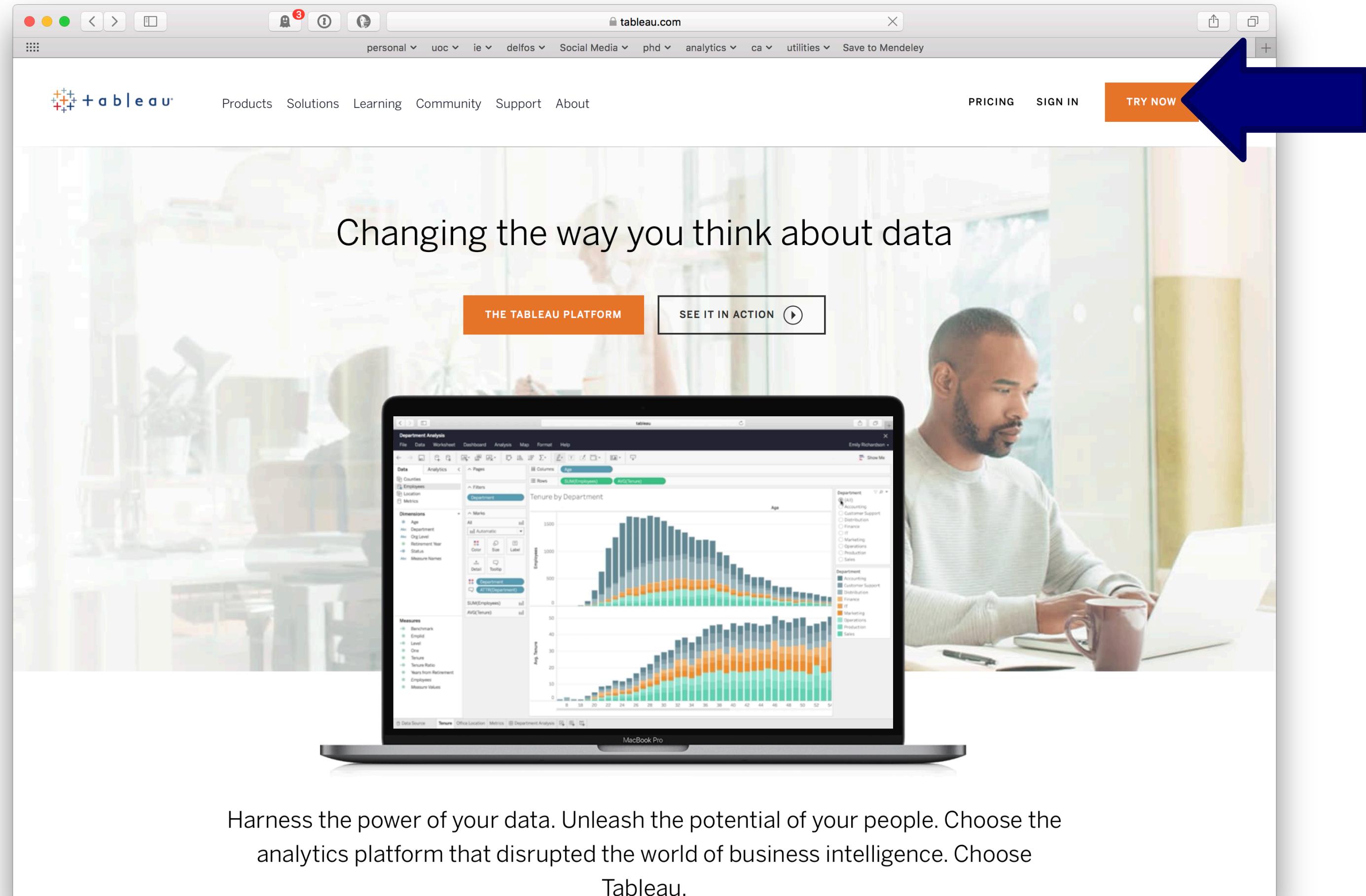
- Windows 7 or newer (64 bit)
- Intel Core i3 or AMD Ryzen 3 Pro or faster
- 4 GB memory
- 2 GM minimum free disk space

Mac

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

If you are not able to install Tableau Desktop, contact the professor

How to install Tableau Prep Builder? (I)

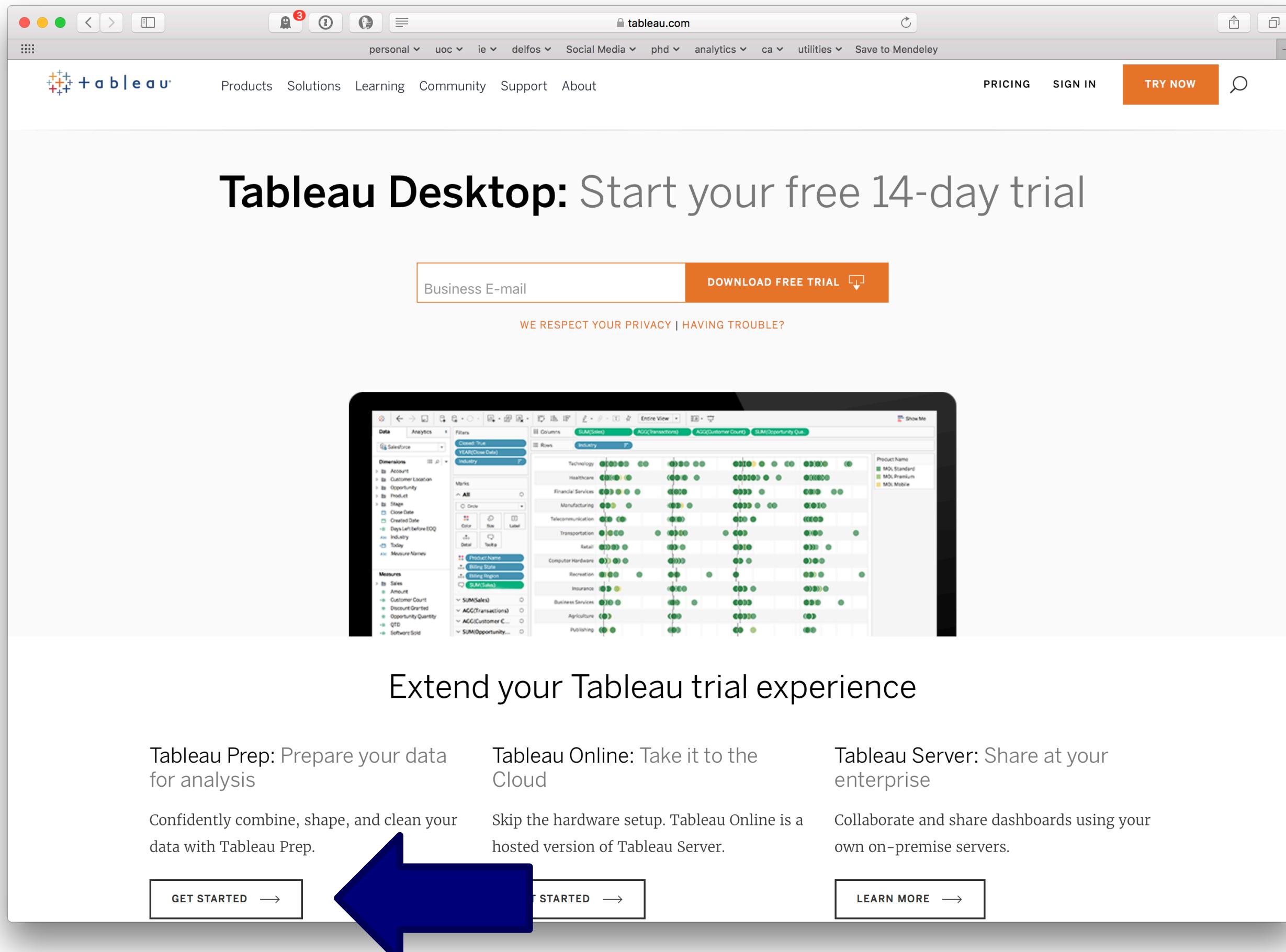


- Go to: <https://www.tableau.com>
- Click **Try Me**

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

Harness the power of your data. Unleash the potential of your people. Choose the analytics platform that disrupted the world of business intelligence. Choose Tableau.

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)

The screenshot shows the Tableau Prep Builder landing page on the tableau.com website. At the top, there's a navigation bar with links like 'personal', 'uoc', 'ie', 'delfos', 'Social Media', 'phd', 'analytics', 'ca', 'utilities', and 'Save to Mendeley'. Below the navigation is the Tableau logo and a search bar. The main heading is 'Tableau Prep: Start your free trial'. There are two prominent orange buttons: 'TRY NOW' and 'START FREE TRIAL'. Below these buttons is a link 'WE RESPECT YOUR PRIVACY | HAVING TROUBLE?'. The lower part of the page displays a screenshot of the Tableau Prep interface. It shows a flow diagram with various nodes like 'Orders (East)', 'Orders (West)', 'Orders (Central)', 'Orders (South)', 'All Orders', 'Orders + Returns', 'Split Customer', 'Quota', 'Pivot Quotas', 'Aggregate', and 'Create 'Annual...''. Below the flow diagram are several data preview tables for fields such as State, Order ID, Segment, Customer ID, Customer Name, and Ship Mode. A 'Changes (2)' section is also visible.

- 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

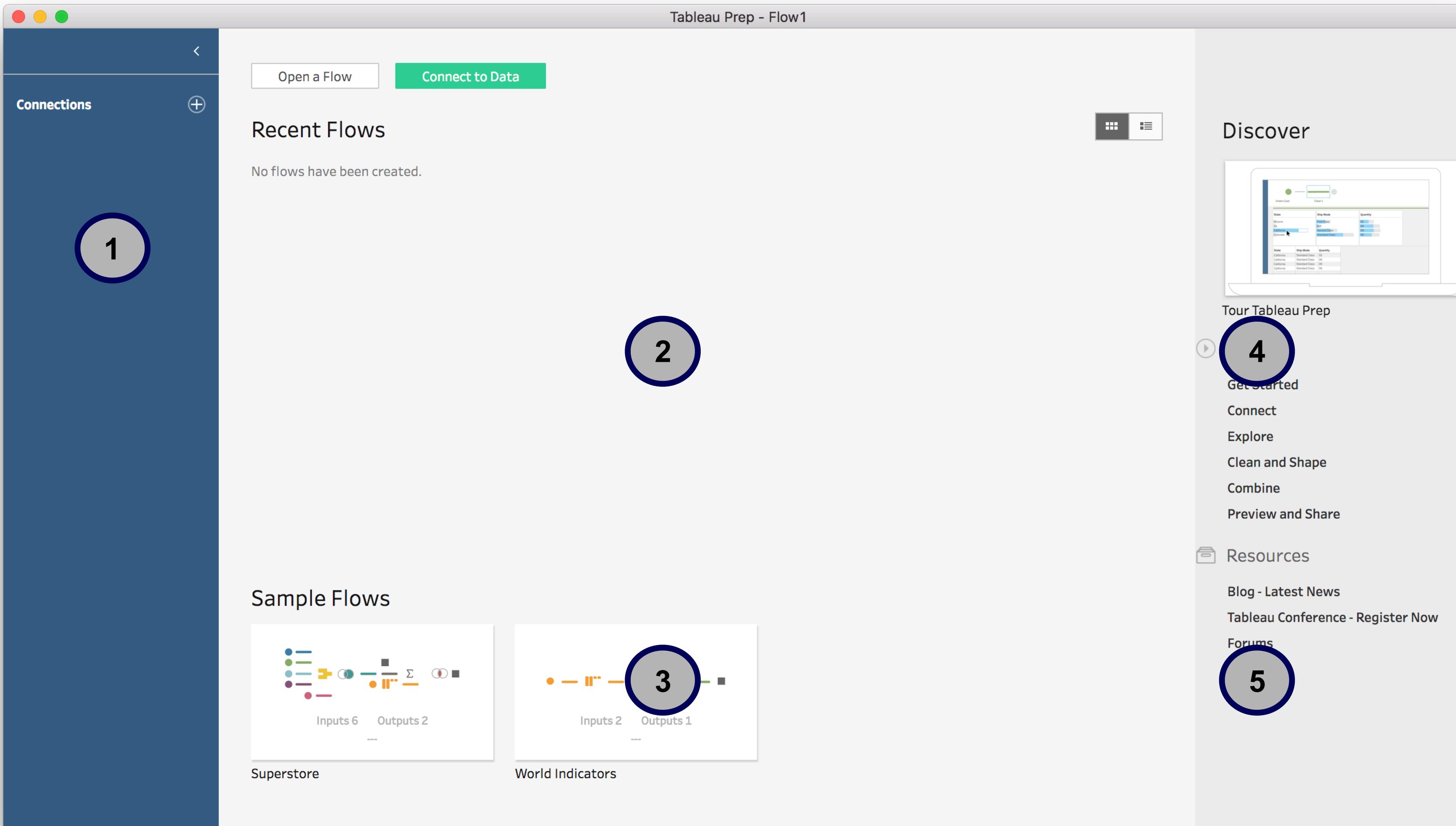
Note: Page may look different.

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,

(Working with) Flows

Tableau Prep Builder Interface

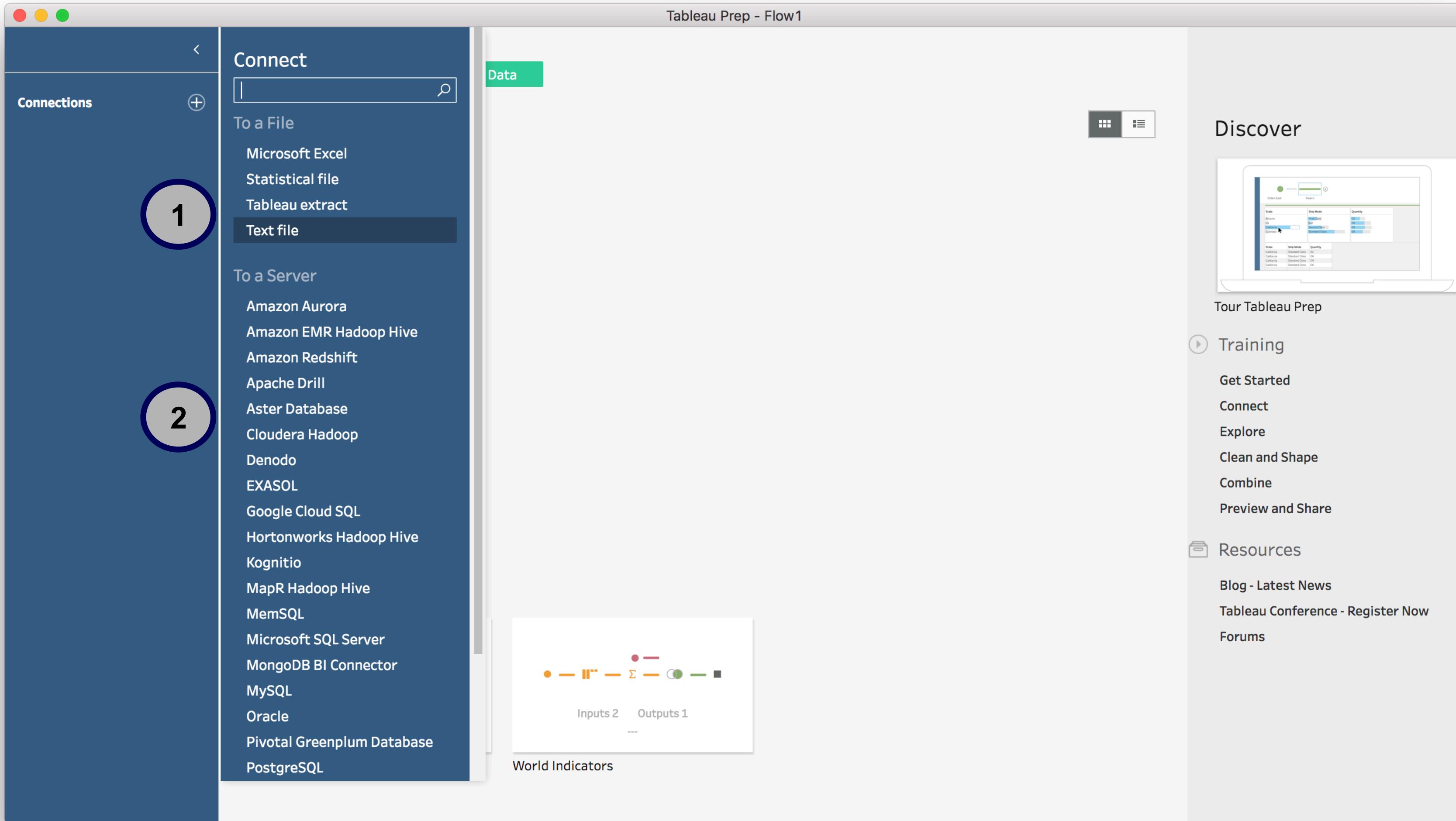
When we open the program, this is the first screen...



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

Data Connectors

There are multiple connectors:



1 Files

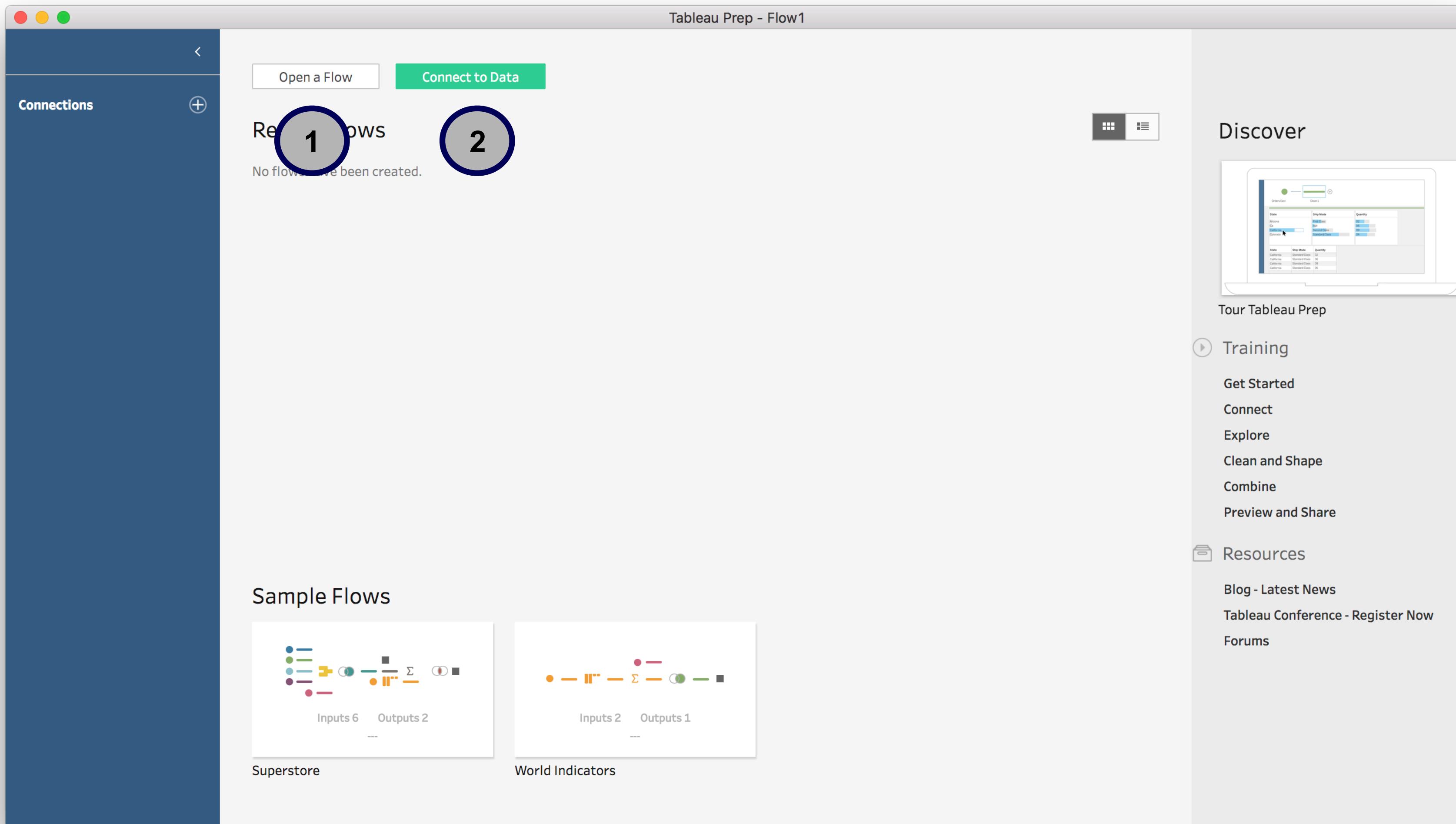
Including excel, text, tableau files and others

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

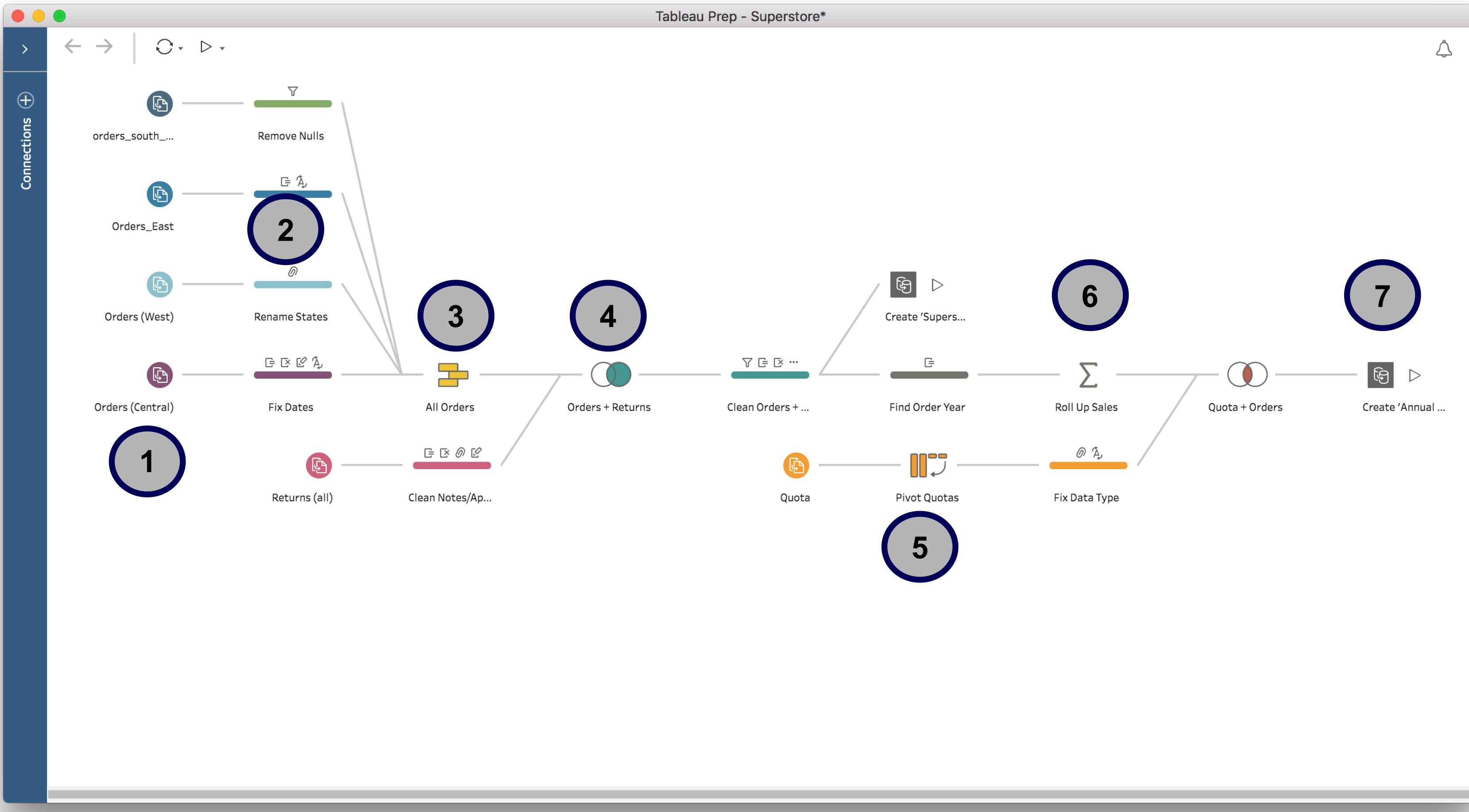
We open a previous flow (to execute or modify it)

2 Data

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

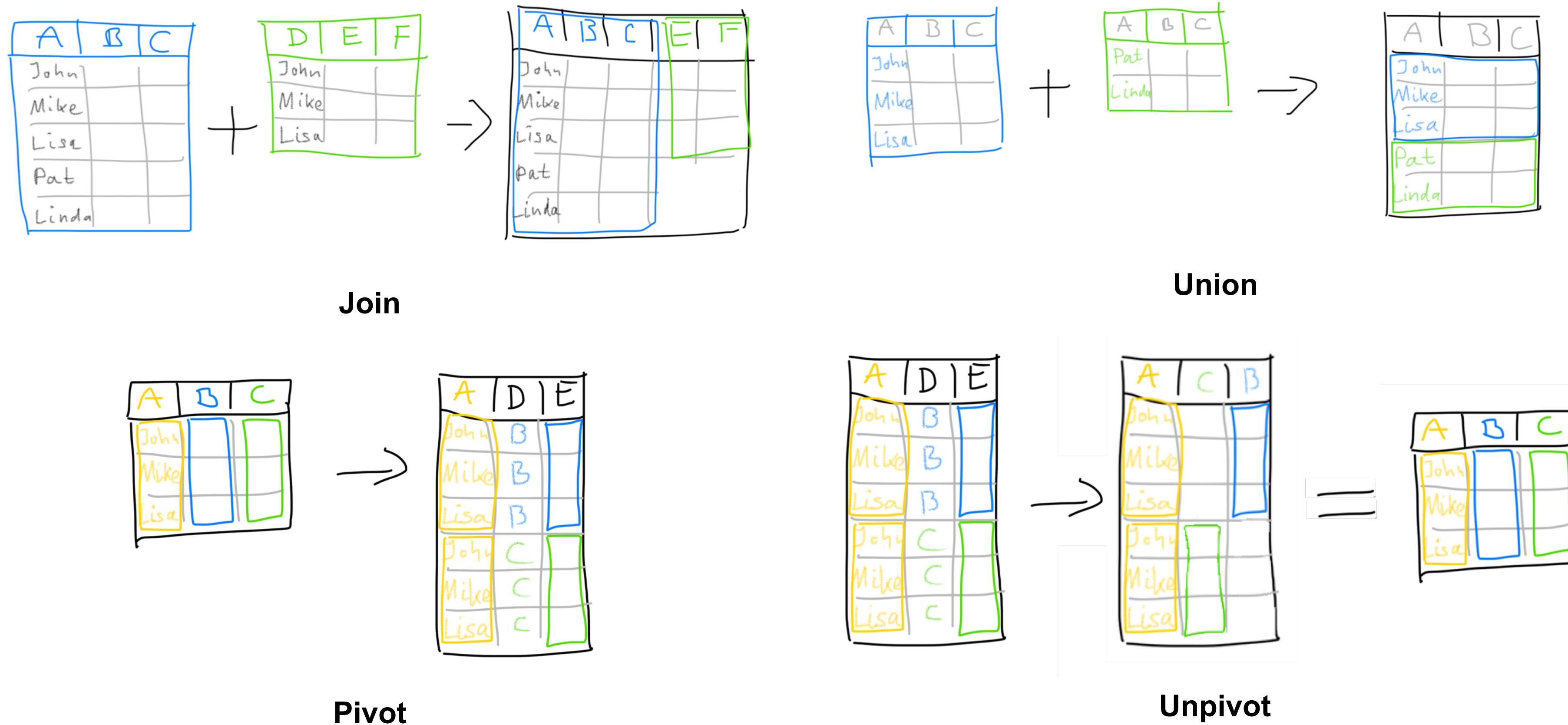
What is a Flow?

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

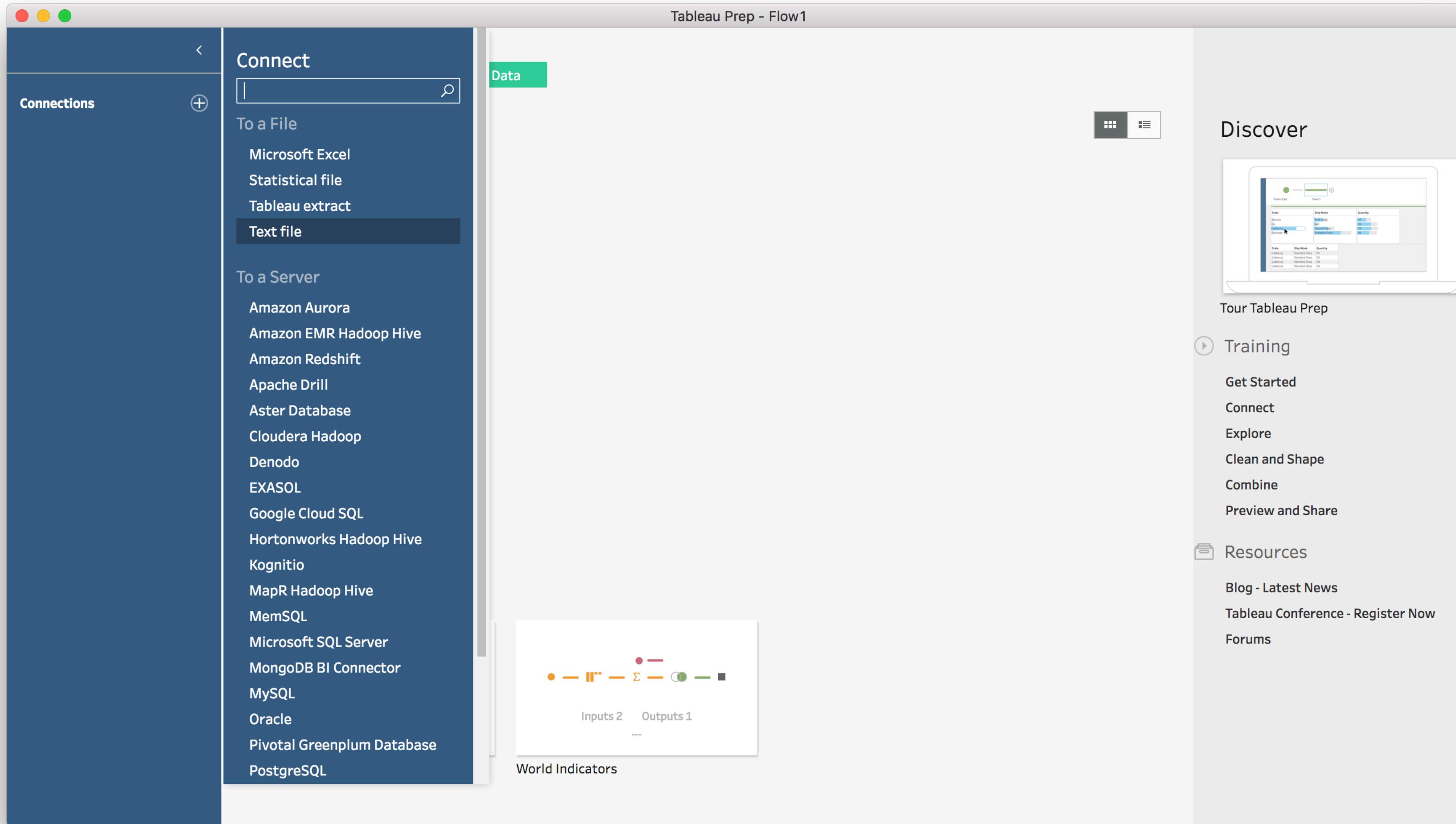


- When we have several data sources, the main options to combine them are: join, union, pivot and unpivot.
- The others steps are related to data cleaning and preparation.

Example

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

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, a 'Data Sample' table titled 'auto_mpg Fields selected: 9 of 9' displays nine fields: mpg, cylinders, displacement, horsepower, weight, acceleration, model_year, origin, and car_name. Each row includes sample values.

- 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)

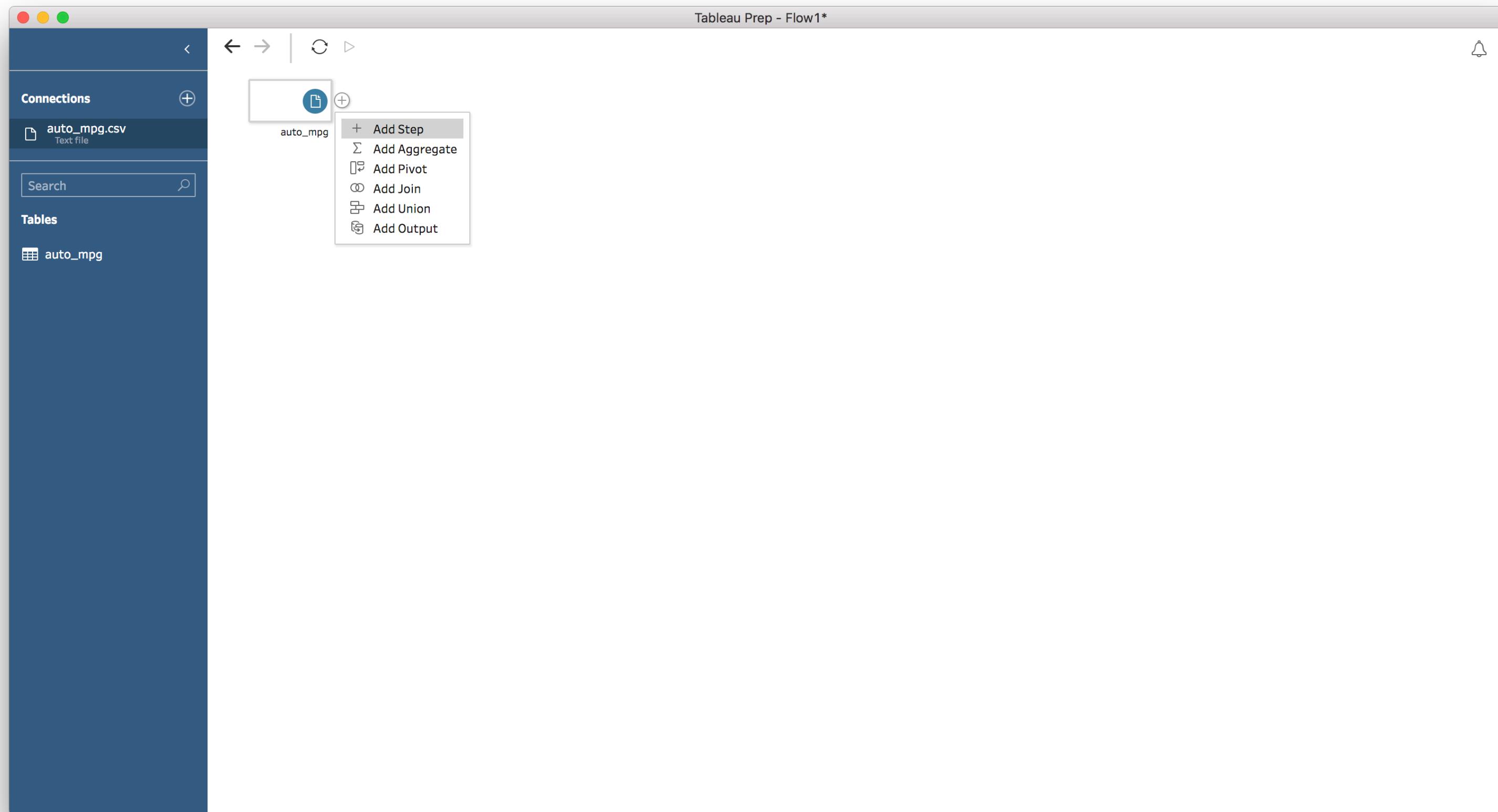
The screenshot shows the Tableau Prep interface. On the left, the 'Connections' pane lists 'auto_mpg.csv' as a Text file. The 'Tables' pane shows 'auto_mpg'. The main area is titled 'Tableau Prep - Flow1*' and contains a 'Data Sample' section for 'auto_mpg'. It displays a table with 9 fields: mpg, cylinders, displacement, horsepower, weight, acceleration, model_year, origin, and car_name. The 'mpg' field is highlighted with a blue border.

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...

- 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.

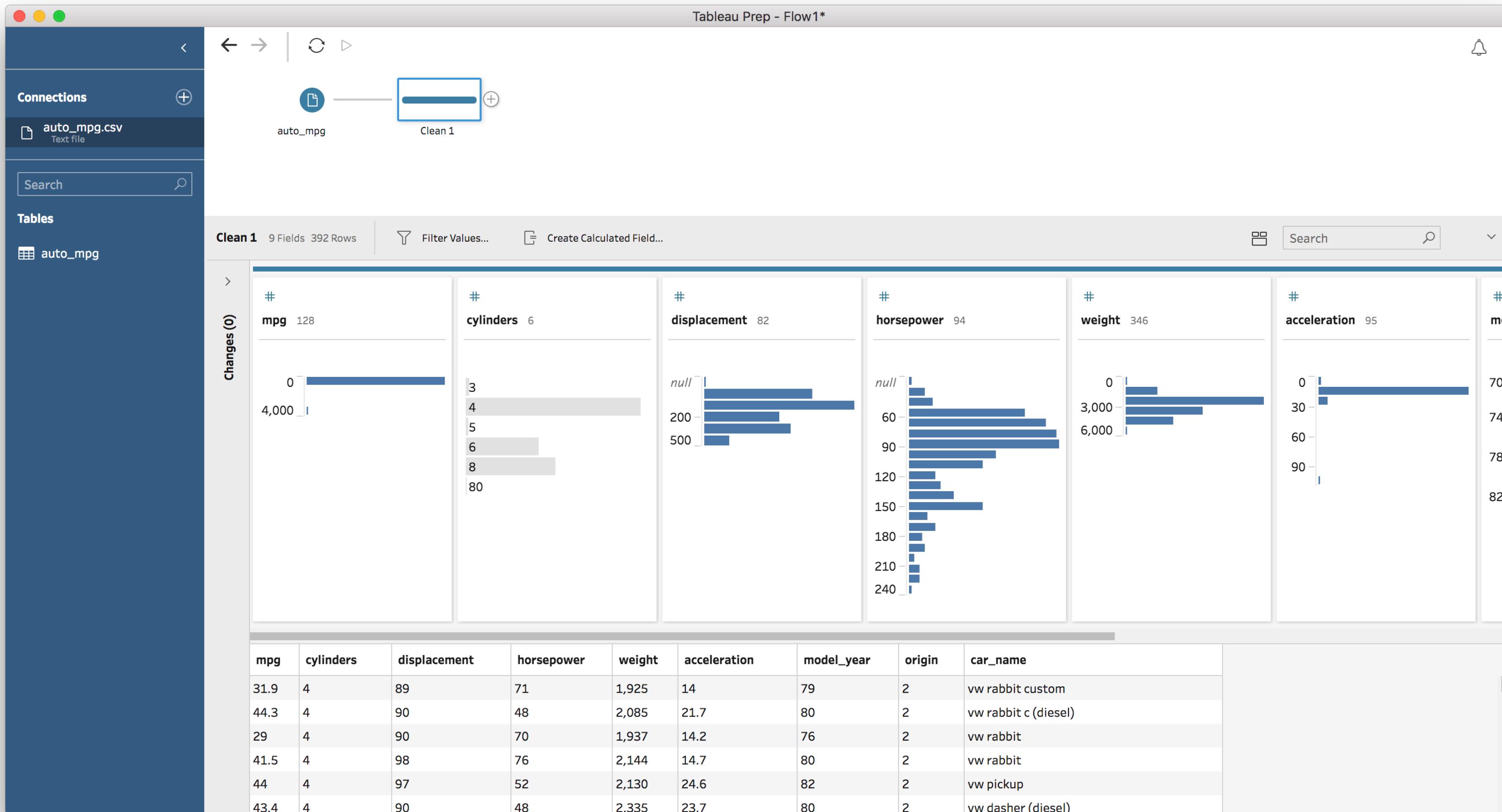


- We have six steps
- **Step:** to apply changes and cleansing functions to our data.
- **Aggregate, pivot, join y union:** to combine the flow of data
- **Output:** to export the flow to a new data set.

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
```

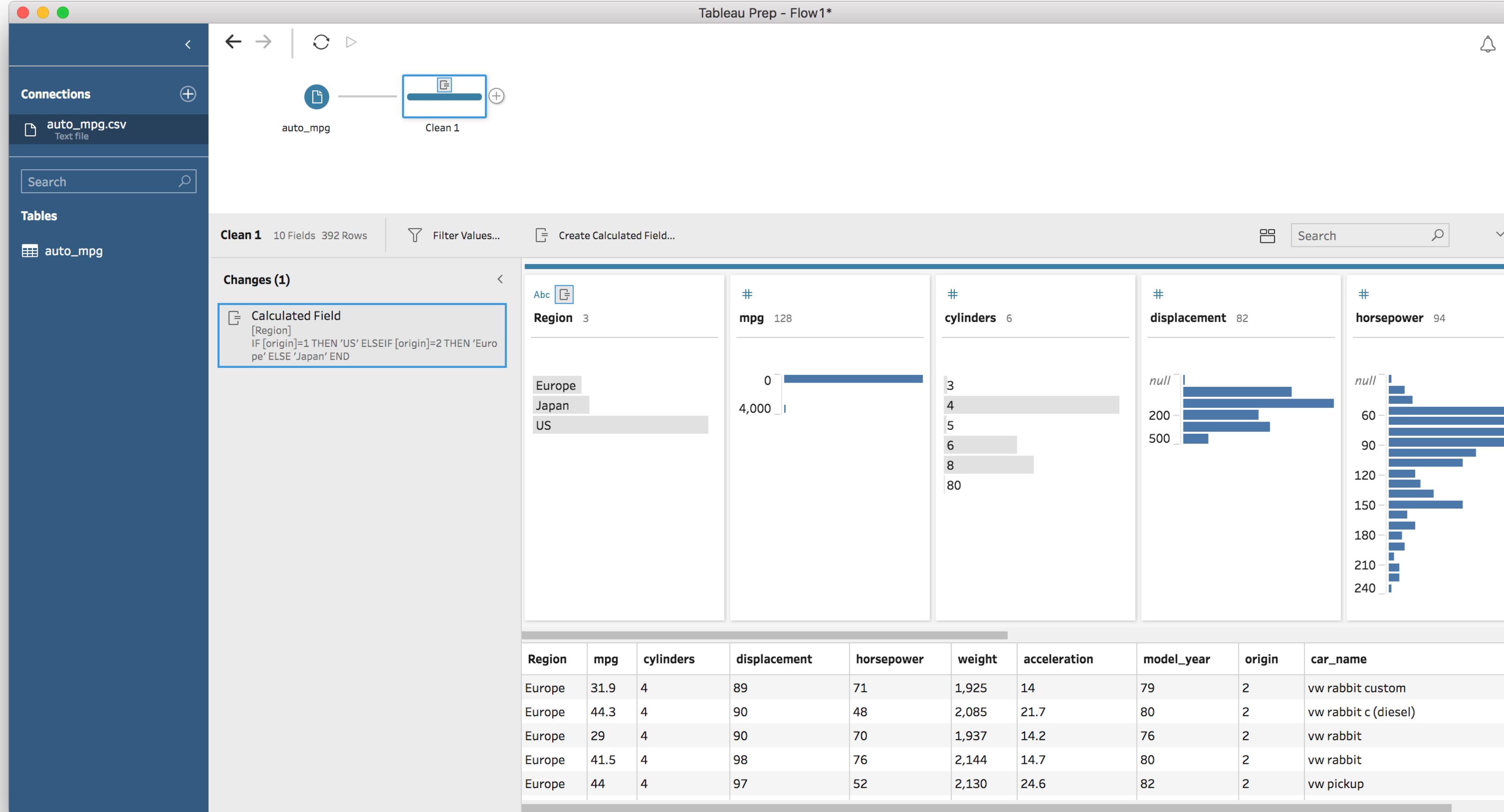
The 'Reference' sidebar shows the 'ABS(number)' function, which returns the absolute value of the given number. Below the formula, a preview of the data shows columns for 'displacement', 'horsepower', 'null', 'model_year', 'origin', and 'car_name'. A small bar chart is also visible.

- 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)

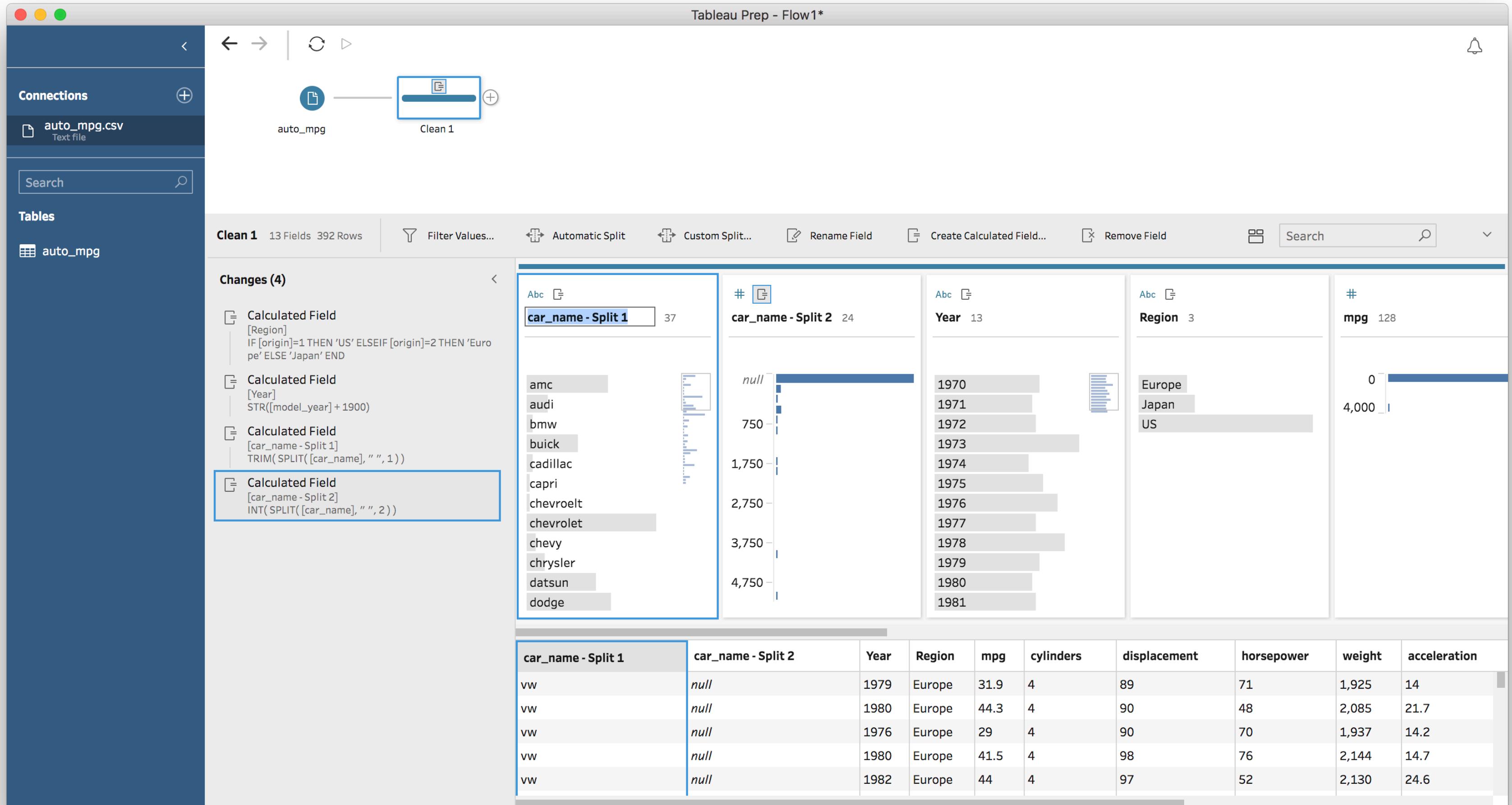
The screenshot shows the Tableau Prep interface with a connection to 'auto_mpg.csv'. A calculated field named 'Year' is being created, defined as `STR([model_year] + 1900)`. The 'Field Name' is set to 'Year'. A tooltip for the `ABS` function is displayed, stating: 'Returns the absolute value of the given number. Example: ABS(-7) = 7'. Below the interface is a preview of the data, showing a histogram of horsepower values ranging from 60 to 240.

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

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

STR([Model Year] + 1900)

Data Preparation (VI)



- 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 'Clean 1'. The flow starts with a connection to 'auto_mpg.csv' and ends with a step labeled 'Clean 1'. The 'Changes (7)' pane on the left lists the following 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]

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 'Flow 1'. The flow consists of a connection to 'auto_mpg.csv' and a step named 'Clean 1'. The 'Changes' pane on the left lists 9 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]
- 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 categories: Europe, Japan, US.
- mpg**: 128 rows with a distribution from 0 to 4,000.
- cylinders**: 6 categories: 3, 4, 5, 6, 8, 80.

A tooltip for the 'brand' column indicates values 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 'Clean 1' stage preview shows four tables: 'car_brand' (29 rows), 'brand' (37 rows), 'year' (13 rows), and 'region' (3 rows). The 'mpg' table preview shows 128 rows with a range from 0 to 4,000. The final output table has columns: car_brand, brand, year, region, mpg, cylinders, displacement, horsepower, weight, acceleration, model_year, origin.

- 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*'. The flow starts with a connection to 'auto_mpg.csv' (Text file) and a step labeled 'Clean 1'. A context menu is open over 'Clean 1' with options like 'Add Step', 'Add Aggregate', etc. Below the flow, the 'Changes (15)' pane is expanded, listing various data transformations applied to fields such as 'car_brand', 'year', 'region', 'mpg', and 'cylinders'. The main workspace displays five preview panes for 'car_brand', 'year', 'region', 'mpg', and 'cylinders', and a data grid at the bottom showing the cleaned data 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 modal dialog box is open over the workspace, titled 'Tableau Prep - Flow1*'. It contains fields for 'Save As:' (set to 'auto_mpg.hyper'), 'Tags:', and 'Where:' (set to 'Datasources'). Below these are 'Cancel' and 'Accept' buttons. The 'Output' section of the main interface shows 12 fields and a preview of the data, which includes columns like car_brand, year, region, mpg, cylinders, displacement, horsepower, weight, acceleration, model_year, origin, and car_name. The data preview lists various car models from the 1970s and 1980s.

- 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. In the center, a flow diagram consists of three steps: 'auto_mpg' (Text file), 'Clean 1' (represented by a blue icon with a checkmark), and 'Output' (represented by a blue icon with a document). A modal dialog titled 'Tableau Prep - Flow1*' is open over the flow, prompting to 'Save As: auto_mpg.tfl'. Other options in the dialog include 'Tags:', 'Where: Flows', and 'Format: Tableau Flow Files'. Below the flow, the 'Output' step is expanded to show a preview of the data. The preview table has 12 columns: car_brand, year, region, mpg, cylinders, displacement, horsepower, weight, acceleration, model_year, origin, and car_name. The data includes rows for various car models like Volkswagen, Volvo, and Ford from different years and regions.

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



