

LAB 8: VISUALISING TEXT AND DOCUMENT



Table of Contents

Learning Outcome..... 2

Data Preparation for Task 1 2

Task 1: Build a Word Cloud 3

Data Preparation for Task 2 7

Task 2: Remove common words from word cloud 8

Learning Outcome

At the end of this session, learners will be able to:

- Build a word cloud
- Remove common words from word cloud

Data Preparation for Task 1

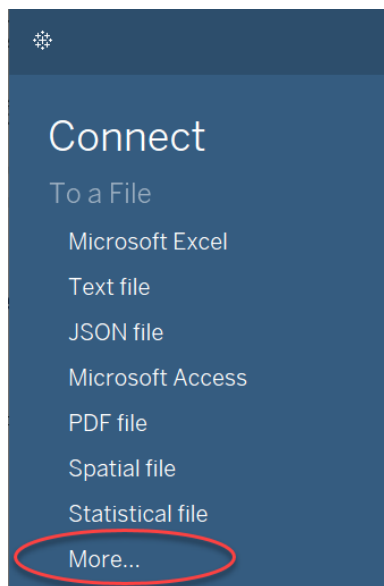
Create a Tableau work book that connect to the **movies data source**.

We are going to use this data source for all the first two tasks of this lab.

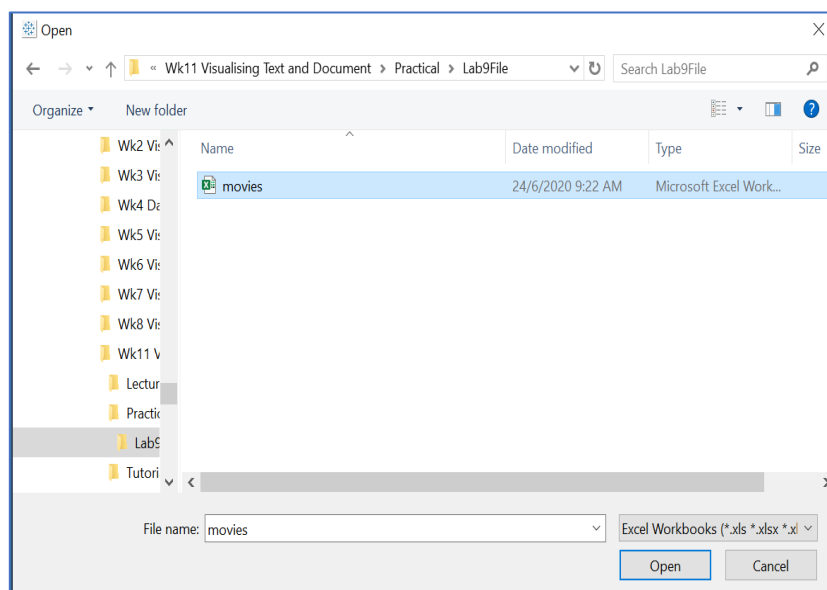
Step 1. Launch Tableau. Under Connect To a File, select **Microsoft Excel**.

Step 2. From the file open window, select the **movies.xls** file.

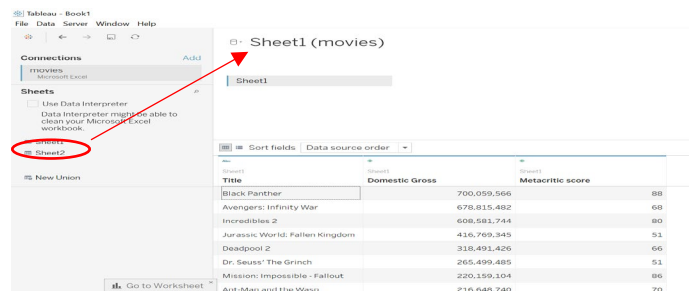
Step 1



Step 2



Step 3: Drag **Sheet1** to the canvas.



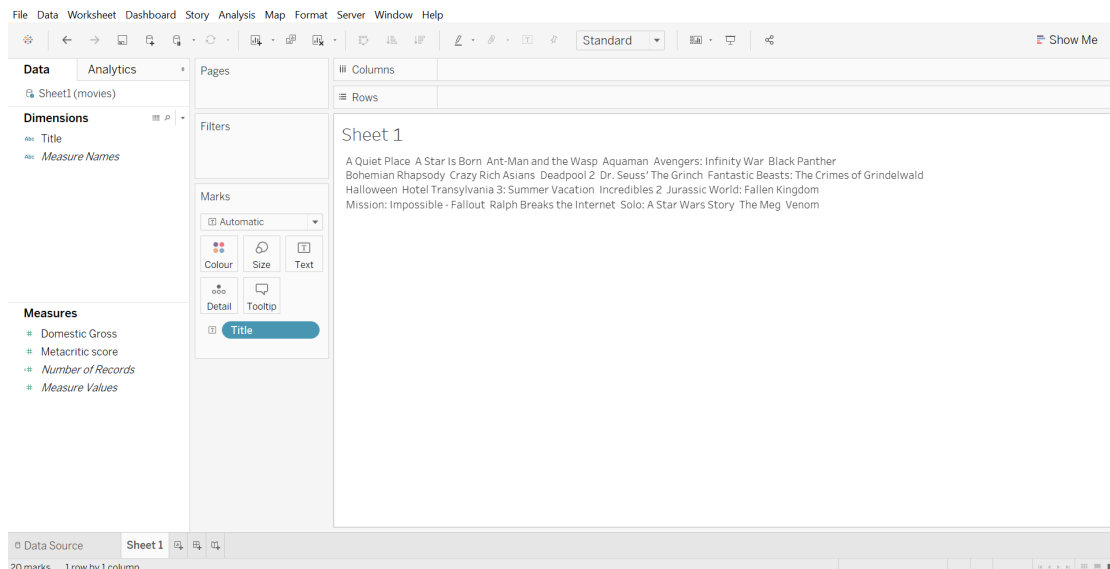
Task 1: Build a Word Cloud

A **Word cloud**, also known as a **Tag cloud**, is a visual representation of text data, typically used to depict keyword metadata (tags) on websites or to visualize free form text[Wikipedia]. Word clouds are a popular type of infographic with the help of which we can show the relative frequency of words in our data.

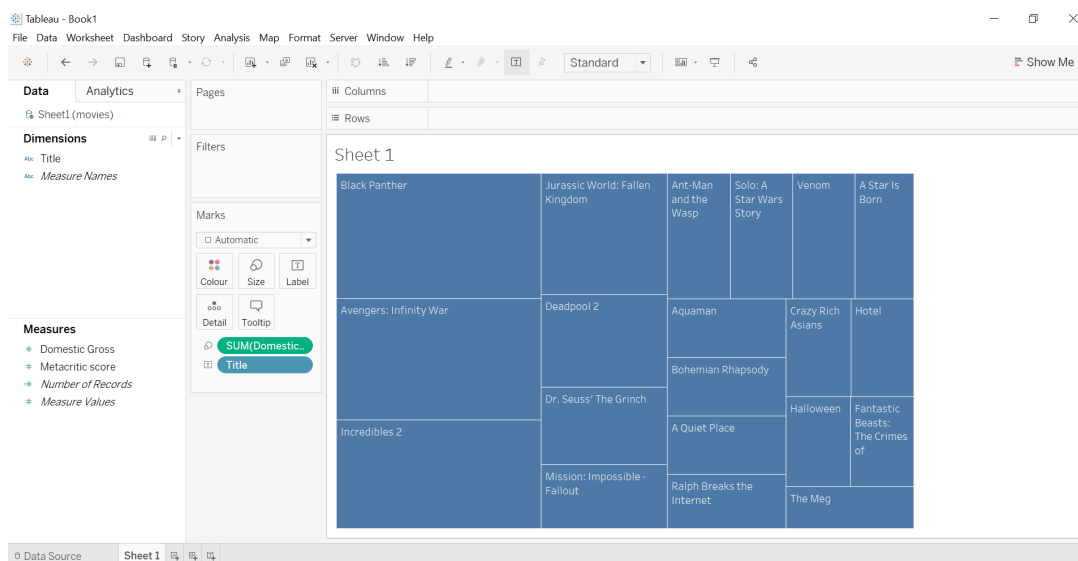
This can be depicted either by the size or the color of the chosen fields in the data. They are a pretty powerful feature to draw attention to your presentation or story

The [data](#) pertains to the top 20 movies of 2018 in the US having been ranked by domestic box office earnings in billions of dollars. The data also contains the **Metacritic** scores for each movie. Metacritic is a website that aggregates reviews of media products including movies.

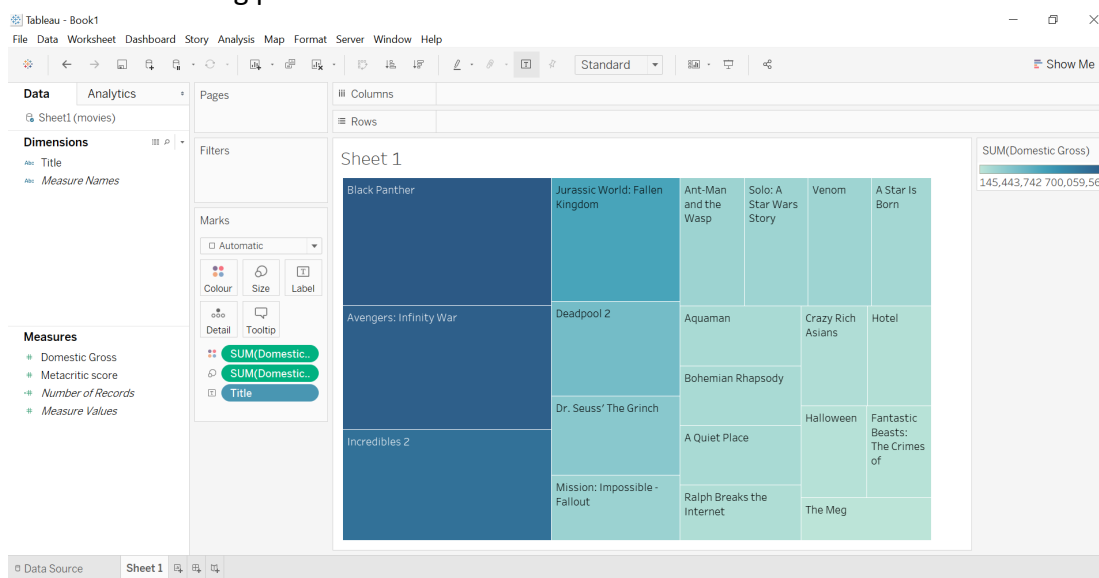
1. Click **Sheet1** tab at the left bottom corner to open the chart designer. Rename it as **Task 1**.
2. Drag the **Title** to **Text** on the **Marks** card for us to know which movie performed well in terms of the Box office collections.



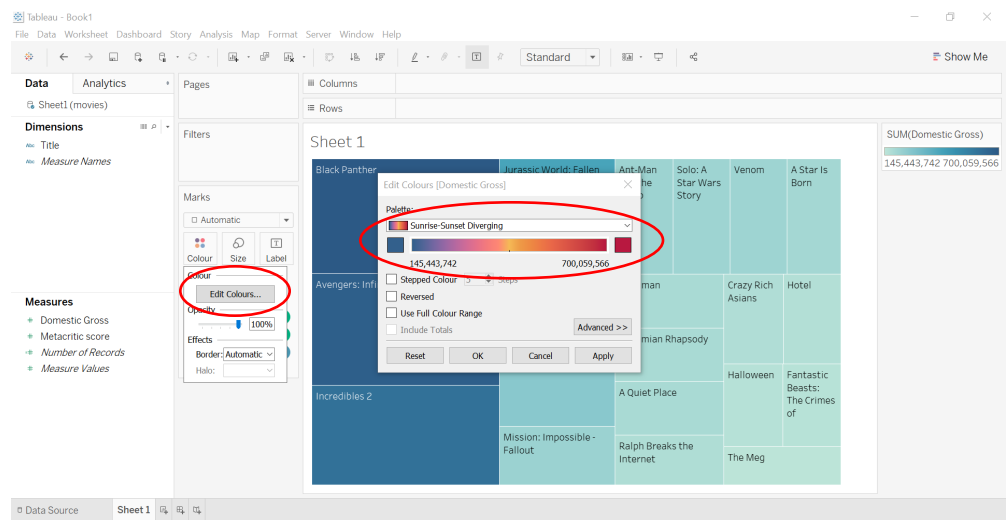
3. Drag the **Domestic Gross** on to the **Size** on the **Marks** card.



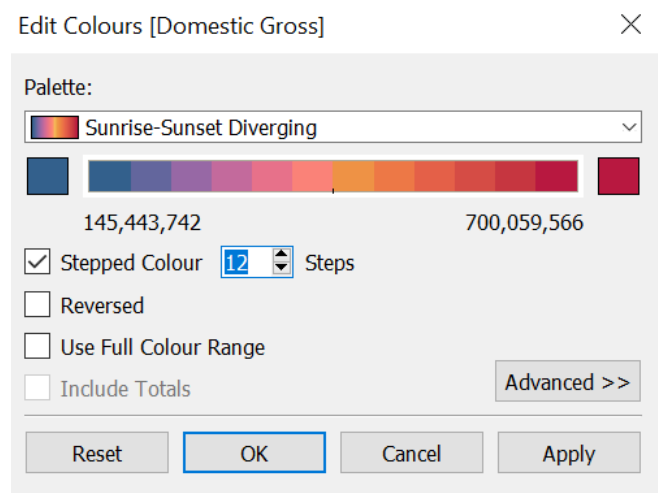
4. Drag the **Domestic Gross** on to the **Color** on the **Marks** card since we want the color to reflect the earning pattern.



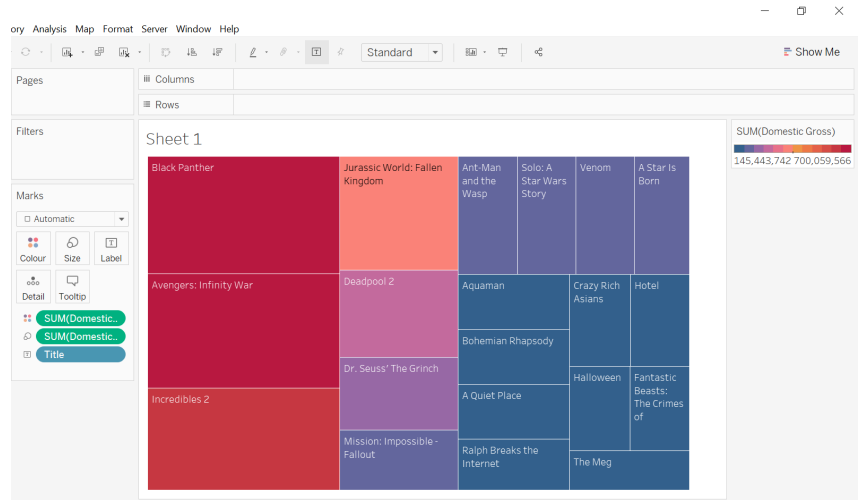
5. Click on the **Edit Color** on **Color** in the **Marks** card. Choose Sunrise-Sunset (Diverging).



6. Select Stepped Color and specify 12 colors for the steps.



7. Below is the look & feel after specifying the colors.



8. Change the **Mark** type from **Automatic** to **Text**.



9. Next, you can change the title, view and background as per your likings and you have your word cloud ready.



Adapted from <https://towardsdatascience.com/word-clouds-in-tableau-quick-easy-e71519cf507a>

Data Preparation for Task 2

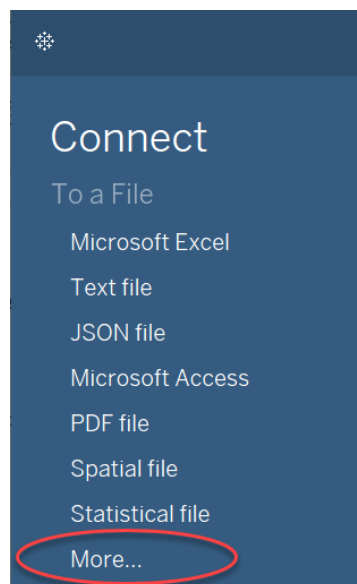
Create a Tableau work book that connect to the **Lyrics Word Count** data source.

We are going to use this data source for all the first two tasks of this lab.

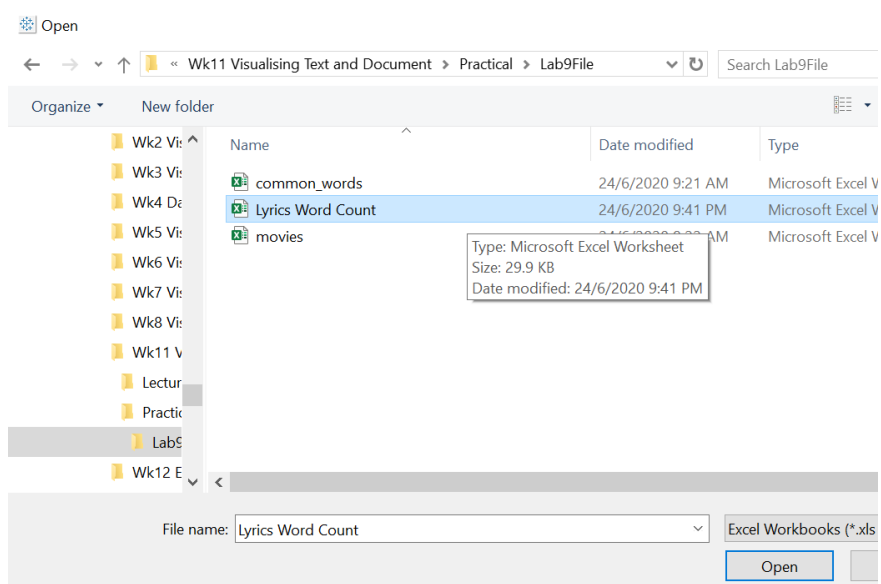
Step 1. Launch Tableau. Under Connect To a File, select **Microsoft Excel**.

Step 2. From the file open window, select the **Lyrics Word Count.xls** file.

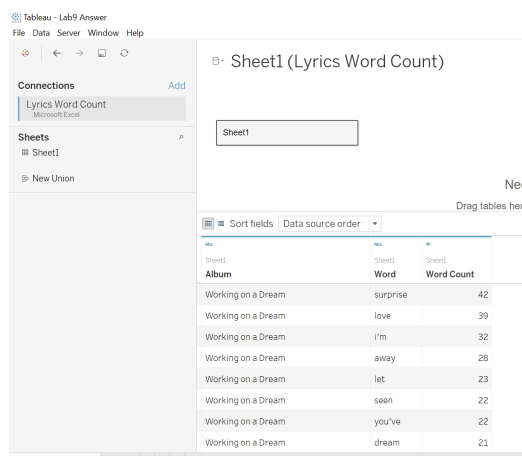
Step 1



Step 2

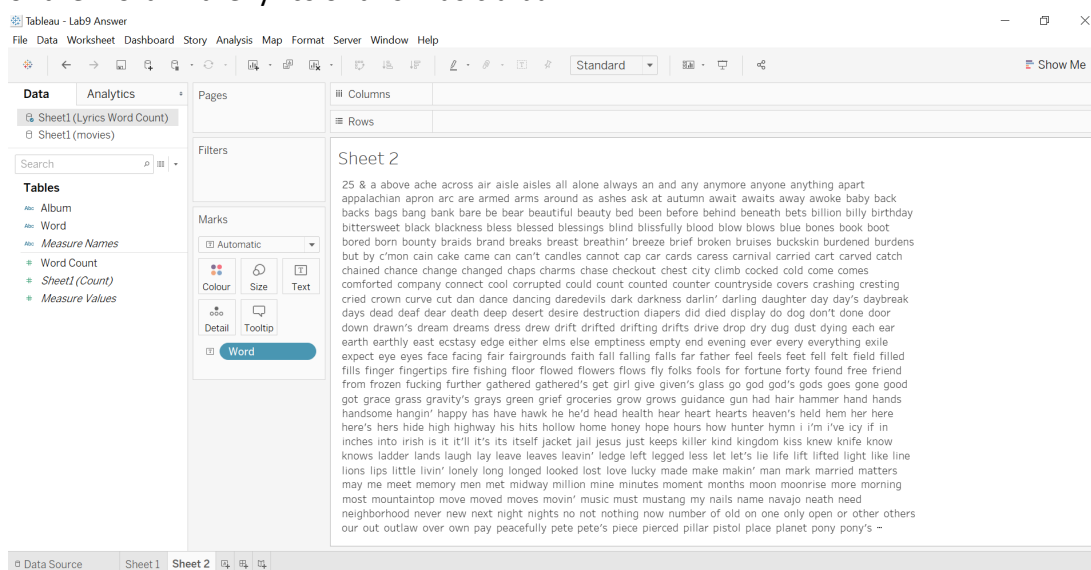


Step 3: The dataset consists of the words and its count of mentions in the album, Working on a Dream.



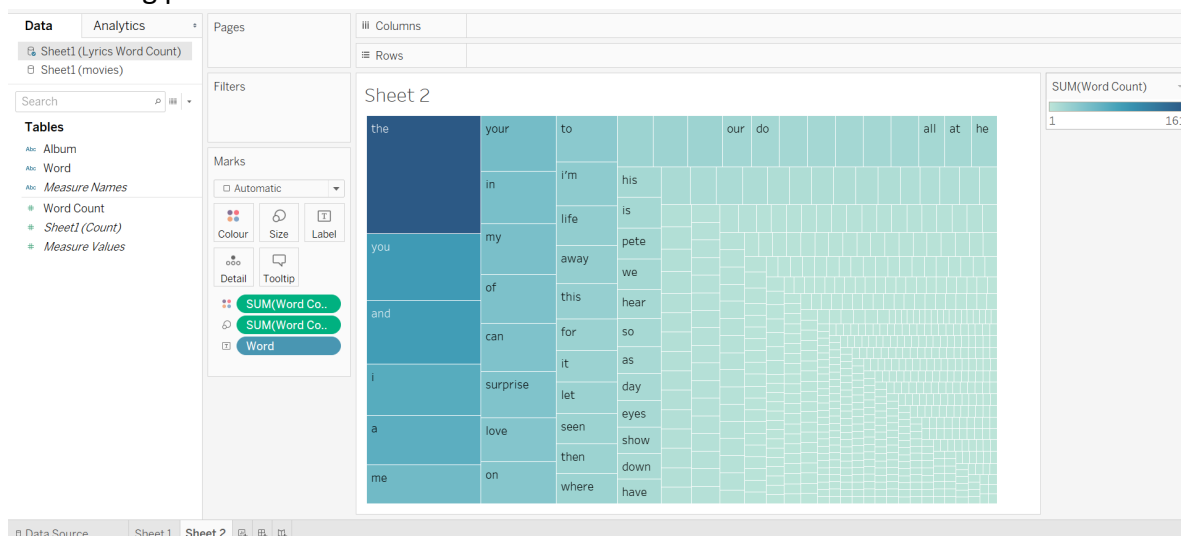
The above examples deal with a simple and refined dataset having limited fields. What if the data contained a paragraph or some passage from a book and we were required to create a word cloud for that. Let's see an example of such an instance.

1. Click **Sheet2** tab at the left bottom corner to open the chart designer. Rename it as **Task 2**.
2. Drag the **Word to Text** on the **Marks** card for us to know which the number of mentions of the word in the lyrics of this music album.

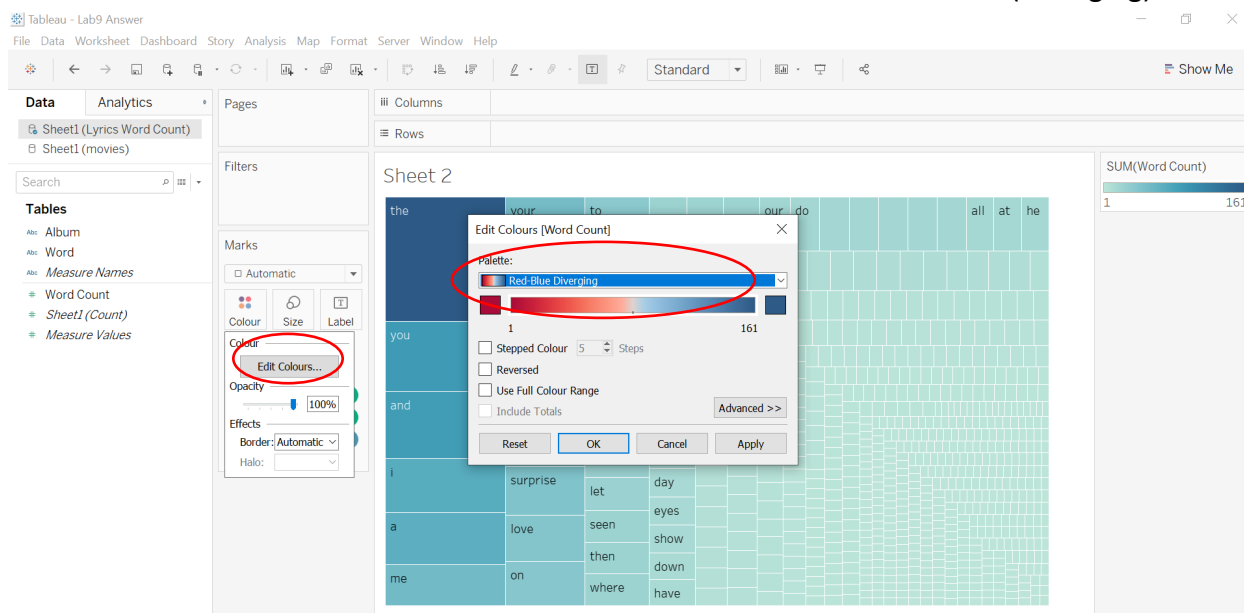


- [illegible]

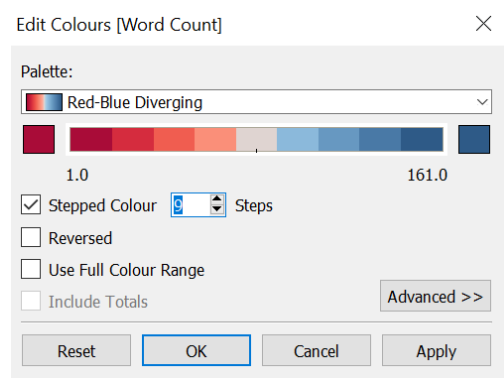
4. Drag the **Word Count** on to the **Color** on the **Marks** card since we want the color to reflect the earning pattern.



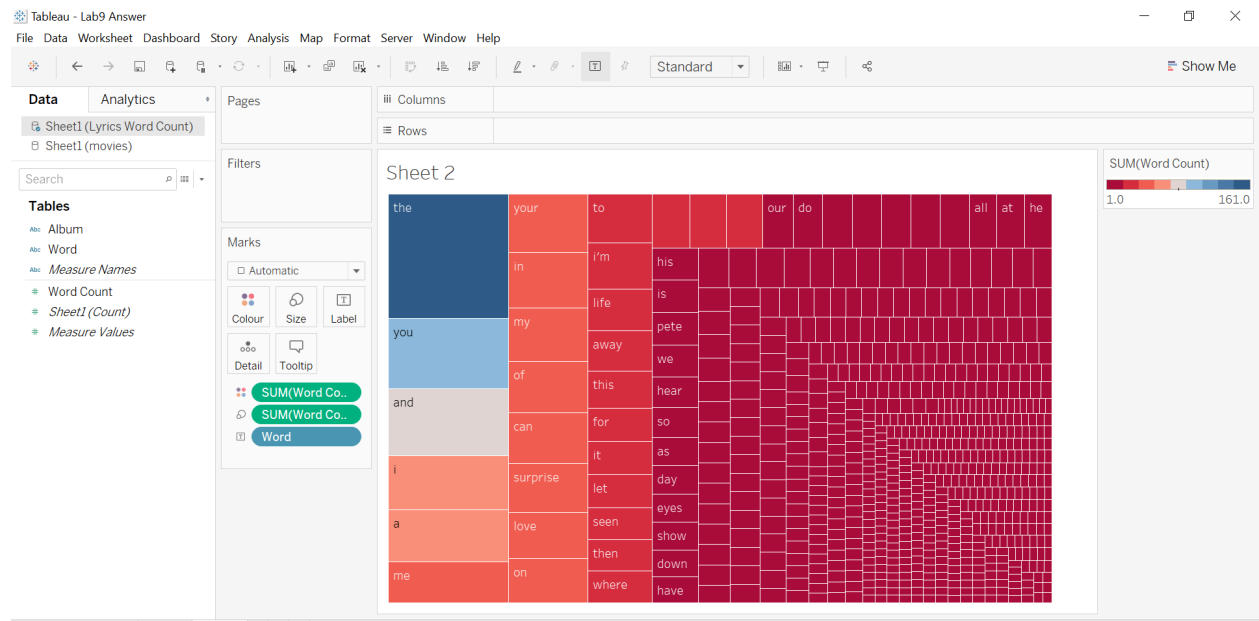
5. Click on the **Edit Color** on **Color** in the **Marks** card. Choose Red-Blue (Diverging).



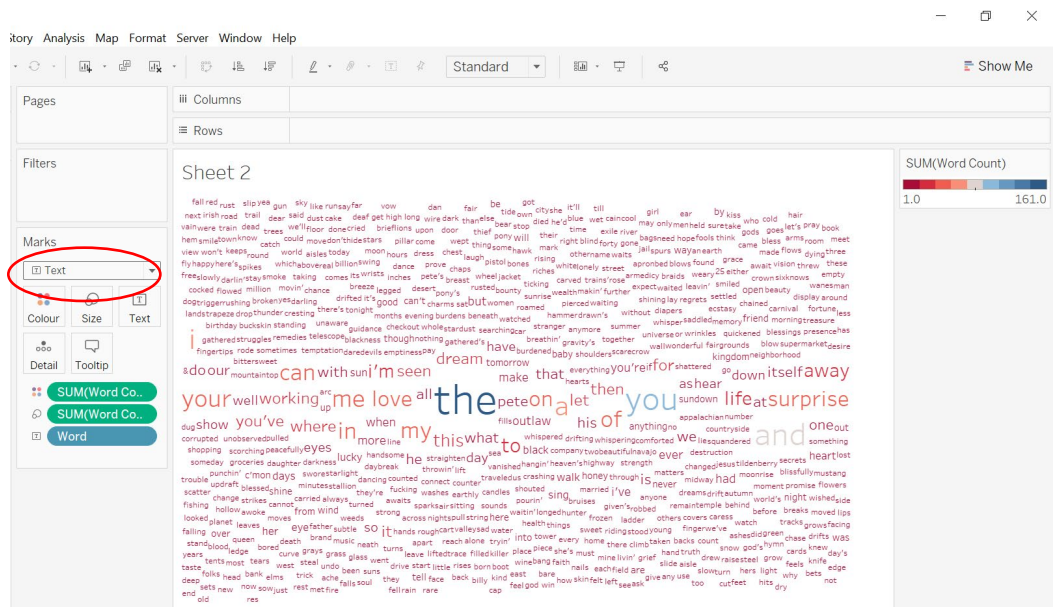
6. Select Stepped Color and specify 9 colors for the steps.



7. Below is the look & feel after specifying the colors.



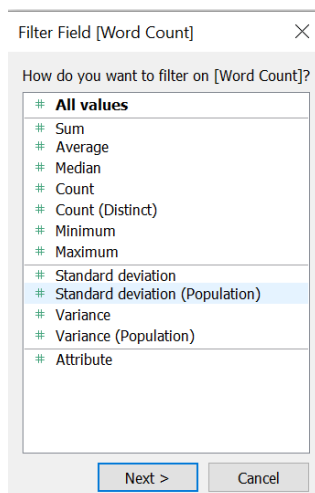
8. Change the **Mark** type from **Automatic** to **Text**.



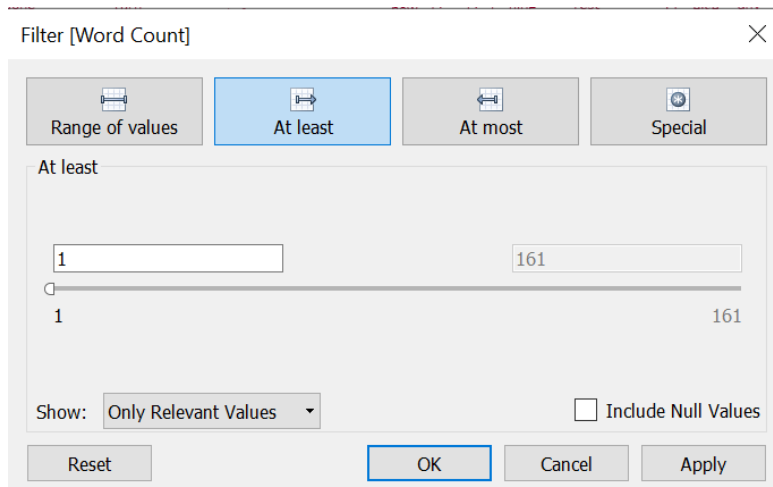
9. Next, you can hide the title, change the view and the background as per your likings and you have your word cloud ready.



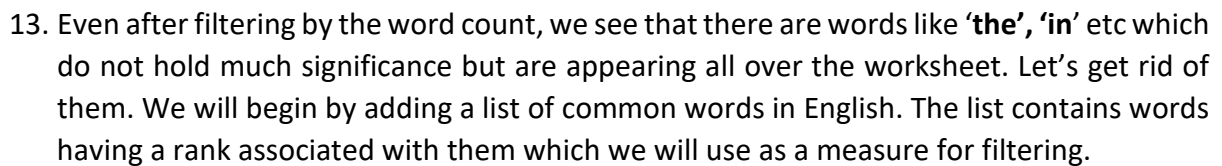
10. Drag and drop word count into Filter. Filter Field window will be displayed. Click on Next button.



11. Select At Least and click OK.

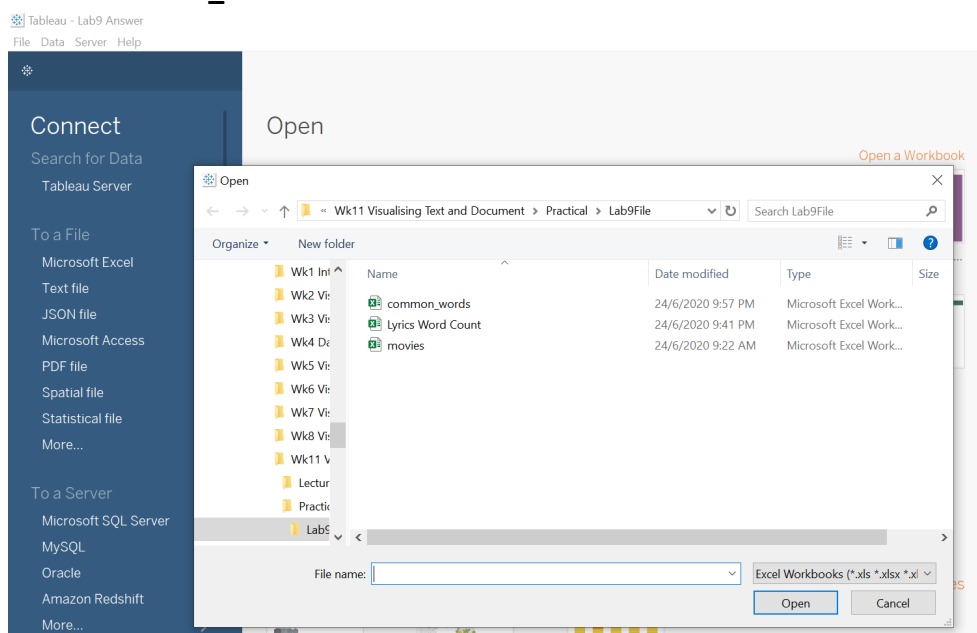


12. Right click the Word Count filter and select Show Filter.

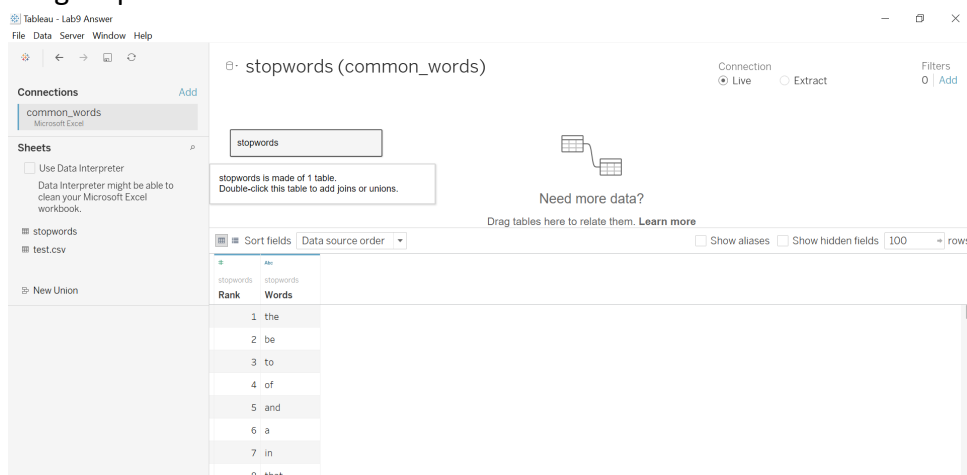


Rank	Word
1	the
2	be
3	to
4	of
5	and
6	a
7	in
8	that
9	have
10	I
11	it
12	for
13	not
14	on
15	with
16	he
17	as
18	you
19	do
20	at
21	this
22	but
23	his
24	by

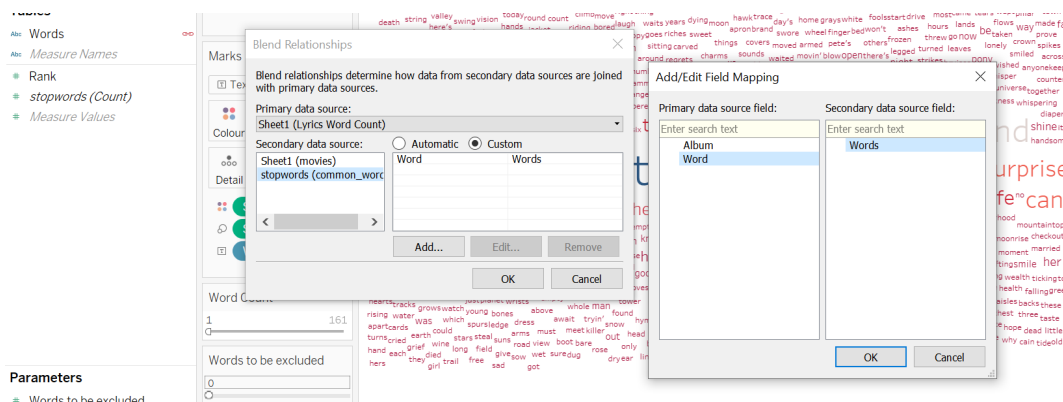
14. Let's add this sheet into our workbook. The two sheets will be blended against the **Words** column since that is common to both sources. Go back to Connect page and select **common_words** excel file.



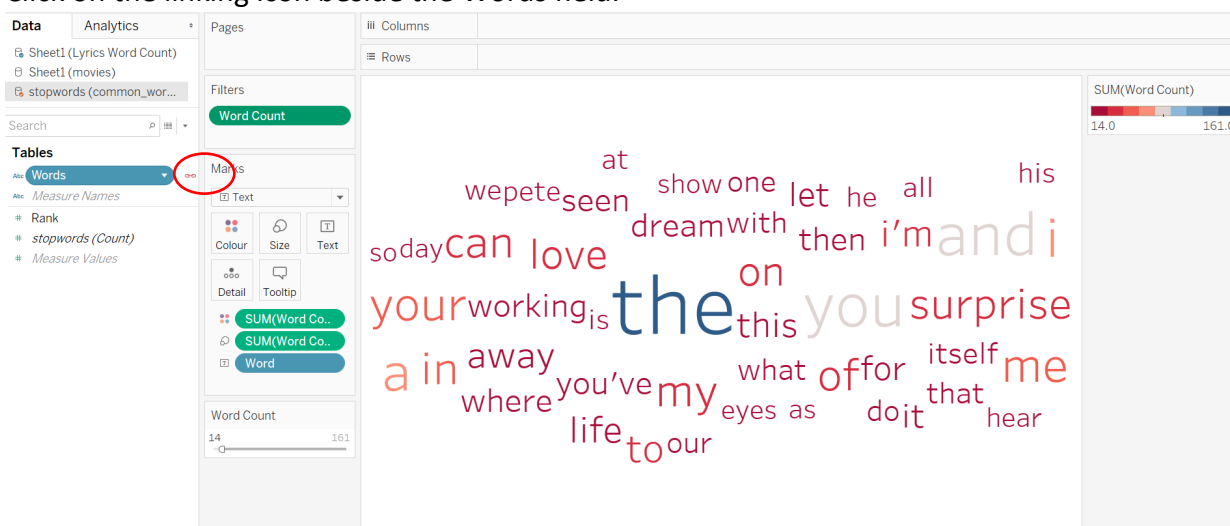
15. Drag stopwords into canvas.



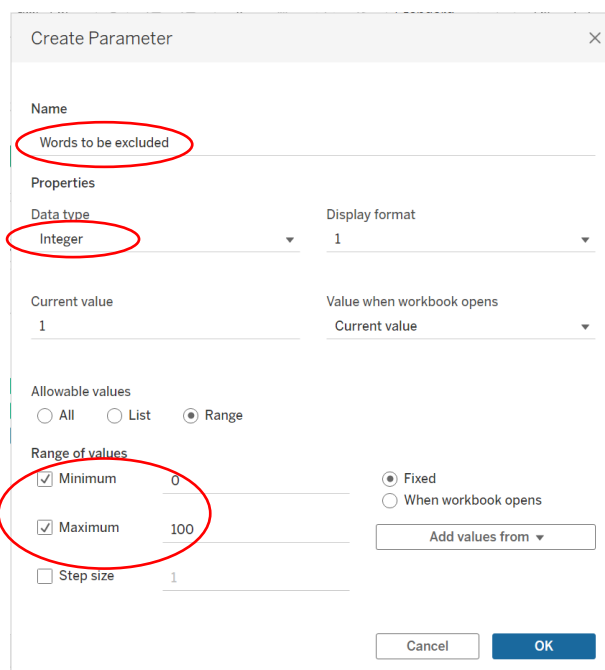
16. Go back Sheet 2 and make sure the stopwords data source is selected. Click on Edit Blend Relationships in Data Menu. Add the field mapping between "word" in Lyrics Word Count and "words" in stopwords.



17. Click on the linking icon beside the Words field.



18. Create a new parameter and name it as **“Words to be excluded”** with the following settings and click OK.



Create Parameter

Name
Words to be excluded

Properties

Data type
Integer

Display format
1

Current value
1

Value when workbook opens
Current value

Allowable values
☐ All
☐ List
☒ Range

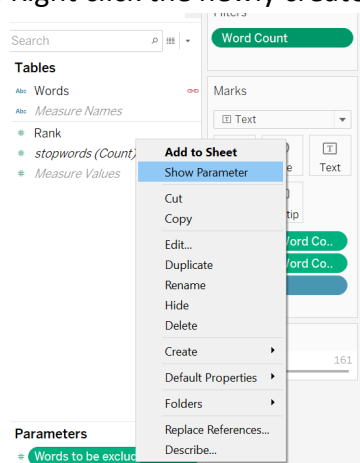
Range of values
☒ Minimum 0
☒ Maximum 100
☐ Step size 1

☒ Fixed
☐ When workbook opens

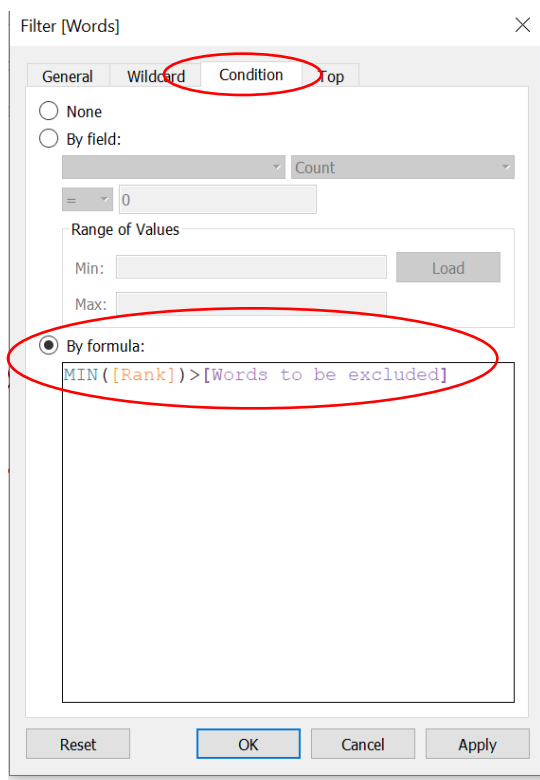
Add values from

Cancel OK

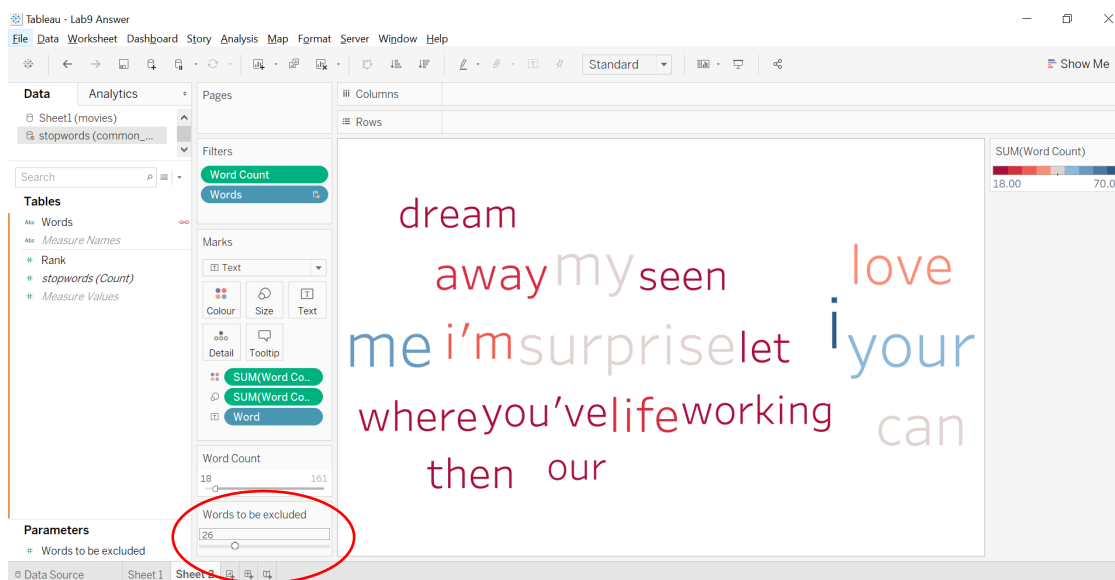
19. Right click the newly created parameter and select Show Parameter.



20. Drag the Words into Filter box and configure the Filter with the following settings.



21. Now you can use the parameter to exclude the common English words. The word cloud should look like the one below.



Adapted from <https://towardsdatascience.com/word-clouds-in-tableau-quick-easy-e71519cf507a>