acaps

HUMANITARIAN ANALYSIS PROGRAM (HAP)

tableau

TRAINING GUIDE

Table of Contents

Introduction

This training will guide you through the development of an operational humanitarian vulnerability dashboard using the desktop tool Tableau.

Scenario

The protection cluster is focussing on identifying the most vulnerable districts within Yemen. Such analysis will support operational teams in planning their humanitarian programming. To give an overview of the most vulnerable areas, the protection cluster have asked you to develop a dashboard highlighting the districts with the most people in need.

Data

The protection cluster have supplied you with two datasets:

- yemen_vulnerability.xlsx – an Excel file containing key metrics of vulnerability.
- Yemen_Admin2.json – a spatial data file of Yemen's admin 2 boundaries.

The dataset is separated by District with each row representing a District. Different geographic levels are represented in the dataset:

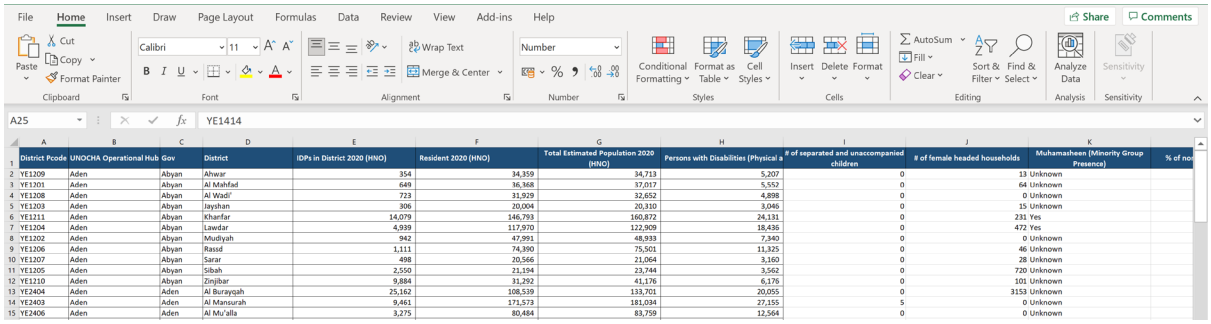
- District – admin level 2 – lowest geographic level of data in this dataset.
- Gov – Governorate admin level 1. Each Governorate contains multiple districts.
- UNOCHA Operational Hub – UNOCHA geographic areas used to support operational planning.
- PCODE – unique geographic code for district. PCODES ensure a consistency in locating geographic areas. This is particularly useful when a geographic area can have multiple spellings.

The dataset contains key vulnerability indicators:

Indicator
IDPs in District 2020 (HNO)
Resident 2020 (HNO)
Total Estimated Population 2020 (HNO)
Persons with Disabilities (Physical and Mental 15%)
# of separated and unaccompanied children
# of female headed households
Muhamasheen (Minority Group Presence)
% of non-functional schools
% of people with access to a regular income
Reduced coping strategies index
Poor to Borderline food consumptions scores (Jan 2019)
% of ind with access to markets
PIN Intercluster

Notice that the data structure is consistent and clean. This flat data structure ensures data quality is high and supports the use of tools such as Tableau or Power BI.

- Row 1 contains the name of each indicator.
- Every other row represents a district. There are no duplicate districts.
- If a indicator contains numerical data, the data does not contains text values.
- The data only contains one table and no graphs are present as these will be created in Tableau.



District Pcode	UNOCHA Operational Hub Gov		District	IDPs in District 2020 (HNO)	Resident 2020 (HNO)	Total Estimated Population 2020 (HNO)	Persons with Disabilities (Physical a	# of separated and unaccompanied children	# of female headed households	Muhamasheen (Minority Group Presence)	% of no
YE1209	Aden	Abyan	Ahwar	354	34,359	34,713	5,207	0	13	Unknown	
YE1201	Aden	Abyan	Al Mahfad	649	36,368	37,017	5,552	0	64	Unknown	
YE1208	Aden	Abyan	Al Wadi'	723	31,929	32,652	4,898	0	0	Unknown	
YE1203	Aden	Abyan	Jaysh an	306	20,004	20,310	3,046	0	15	Unknown	
YE1211	Aden	Abyan	Khanfar	14,079	146,793	160,872	24,131	0	231	Yes	
YE1204	Aden	Abyan	Lawdar	4,939	117,970	122,909	18,436	0	472	Yes	
YE1202	Aden	Abyan	Mudiyah	942	47,991	48,933	7,340	0	0	Unknown	
YE1206	Aden	Abyan	Rasid	1,111	74,390	75,501	11,325	0	46	Unknown	
YE1207	Aden	Abyan	Sarar	498	20,566	21,064	3,160	0	28	Unknown	
YE1205	Aden	Abyan	Sibah	2,550	21,194	23,744	3,562	0	720	Unknown	
YE1210	Aden	Abyan	Zinjibar	9,884	31,292	41,176	6,176	0	101	Unknown	
YE2404	Aden	Aden	Al Buraygah	25,162	108,539	133,701	20,055	0	3153	Unknown	
YE2403	Aden	Aden	Al Mansurah	9,461	171,573	181,034	27,155	5	0	Unknown	
YE2406	Aden	Aden	Al Mu'alla	3,275	80,484	83,759	12,564	0	0	Unknown	

Ensuring data follows a consistent structure supports interoperability between different datasets and allows complex analysis to be developed in a timely, consistent and accurate manner. Similarly, the use of PCODES supports accurate geographic mapping of the data.

For this training you will only be using a few vulnerability metrics. Further details of the data can be found in Appendix 1.

Tableau installation

To install Tableau Desktop, go to https://www.tableau.com/en-gb/support/releases.

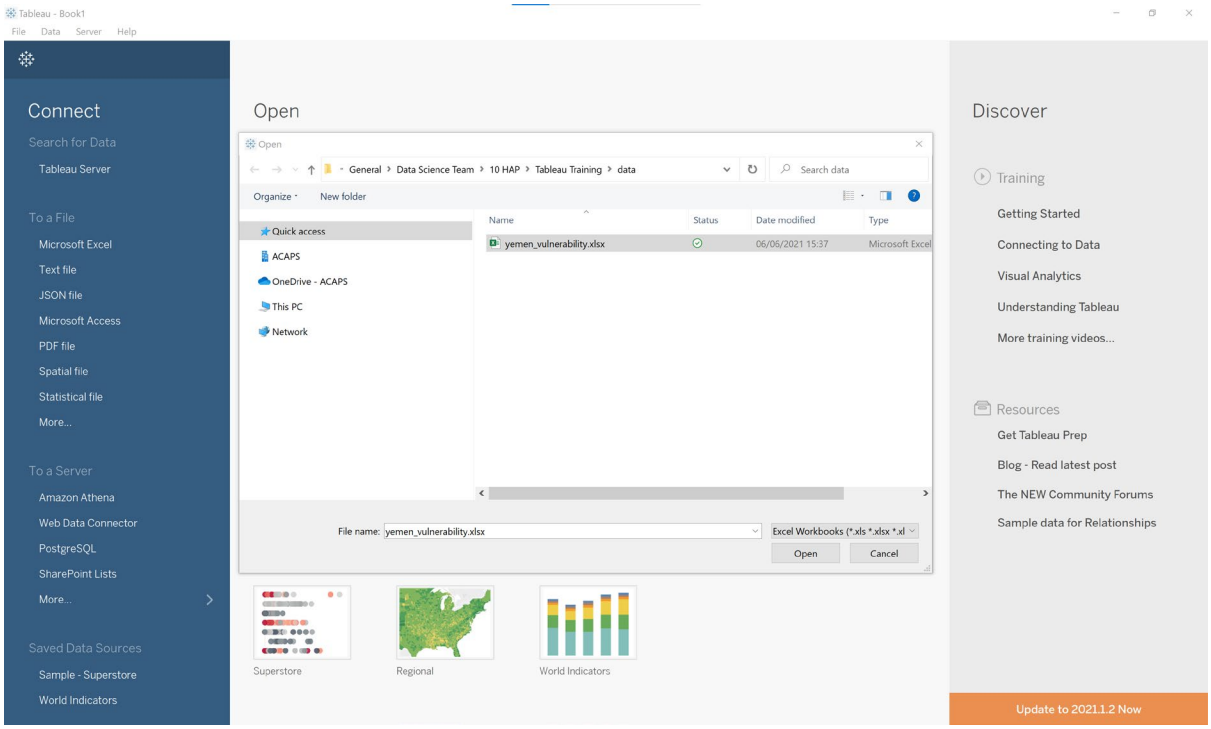
It is also possible to download a free 14 day trial through https://www.tableau.com/products/trial.

Follow the install wizard, then open Tableau Desktop.

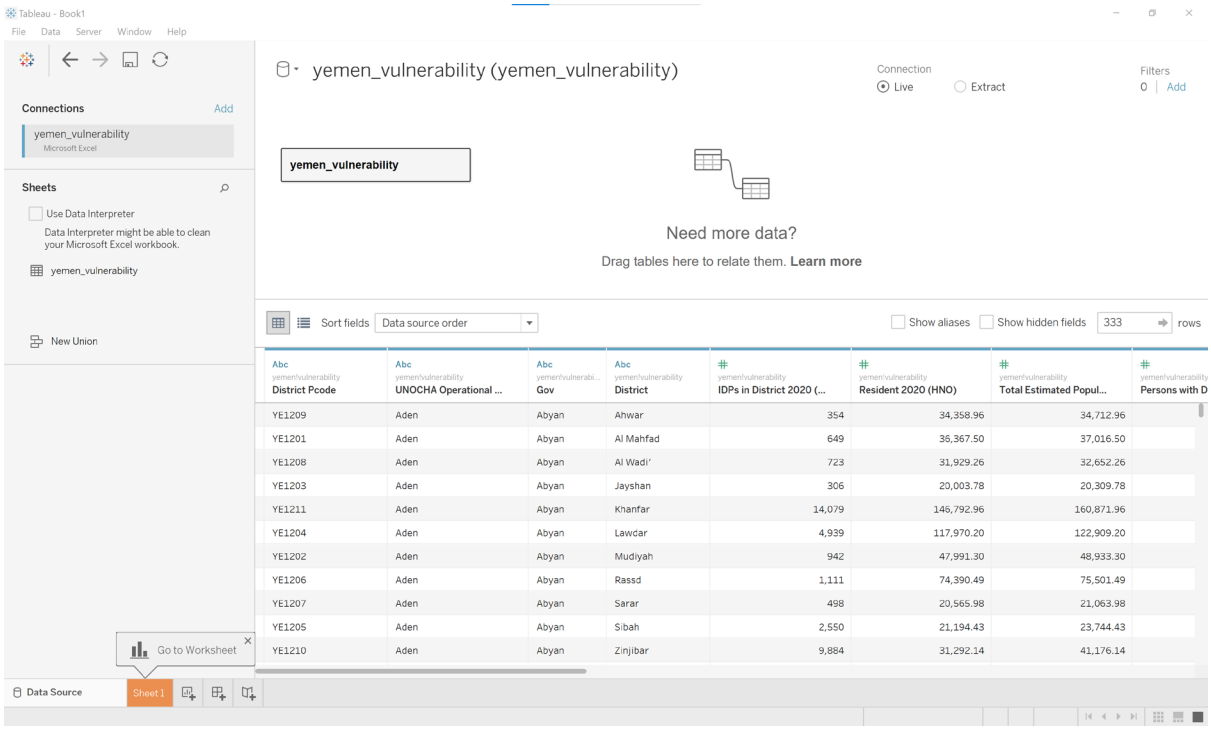
Exercise instructions

1. First we will inspect the data 'yemen_vulnerability.xlsx' in Excel. This dataset contains key vulnerability indicators:

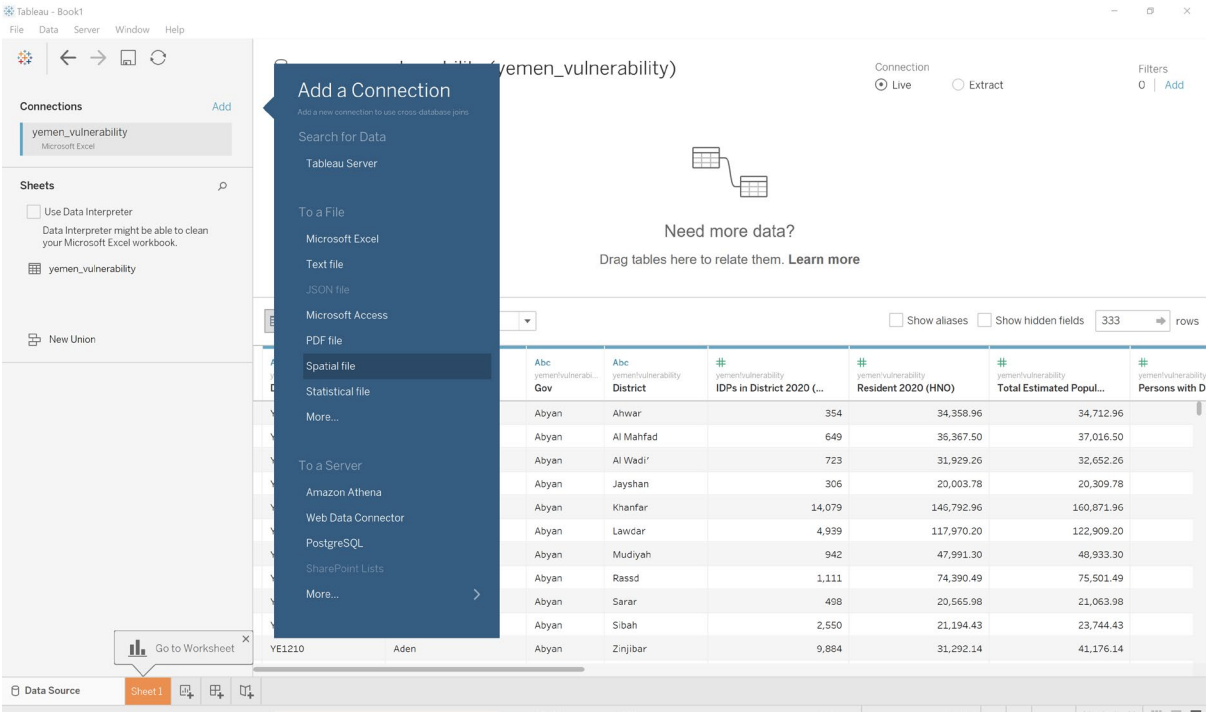
2. Open Tableau Desktop and select Connect -> To a File -> Microsoft Excel.



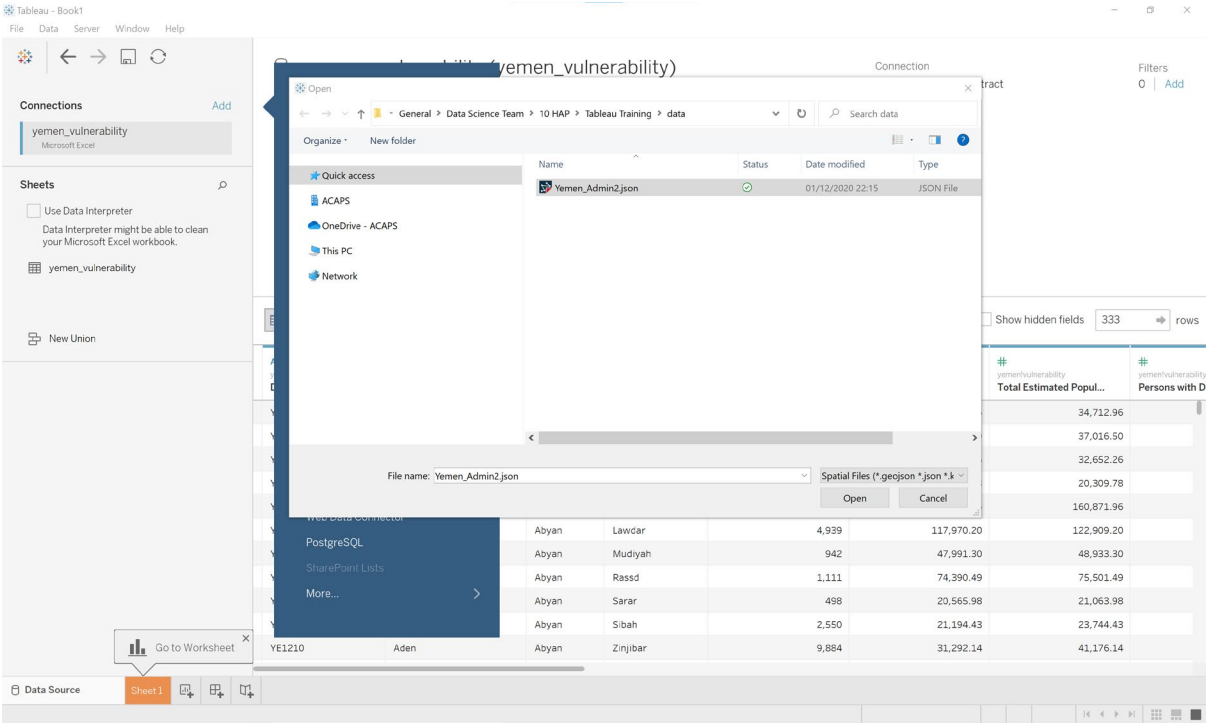
3. Once connected, drag the 'yemen_vulnerability' tab into the data window.



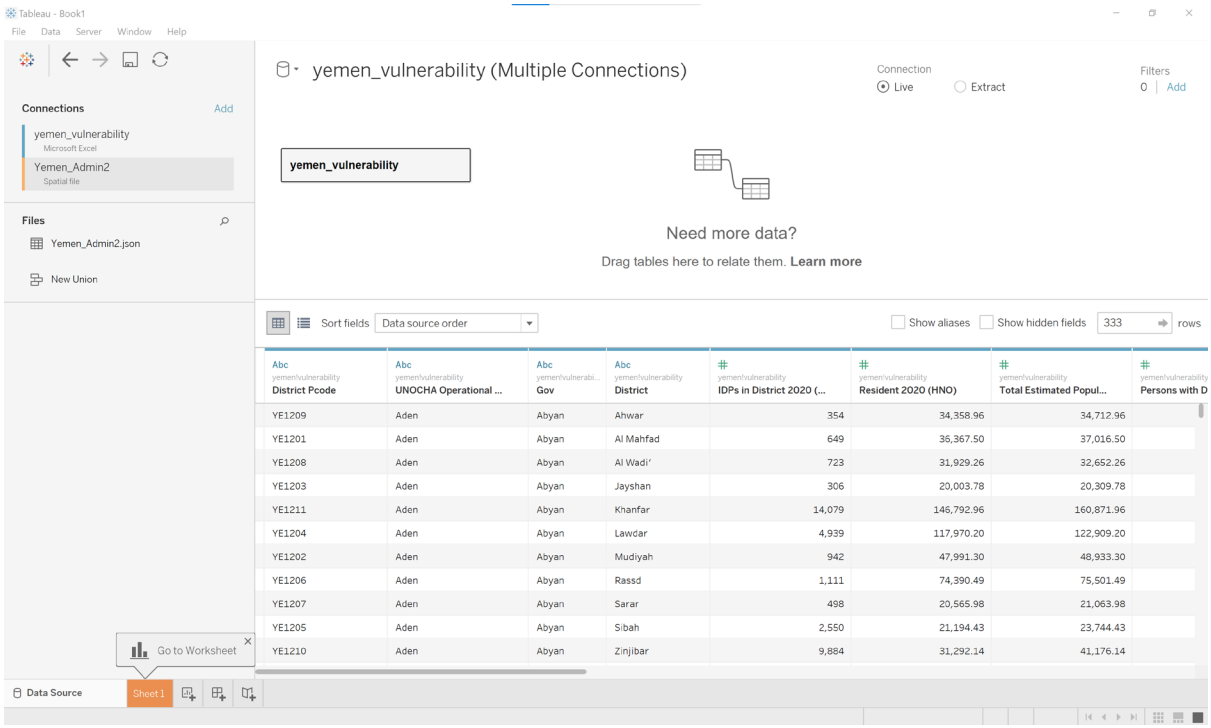
4. Next, add the spatial JSON file of the country boundaries through selecting Add -> Spatial file.



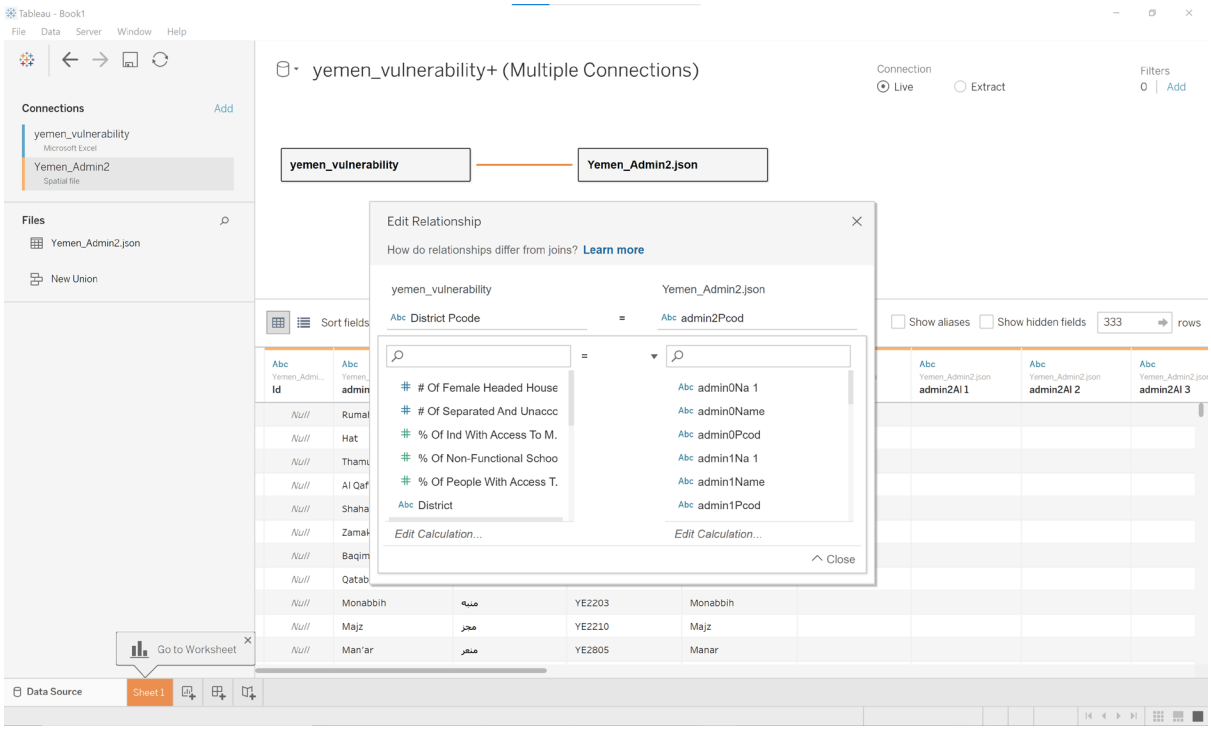
5. Navigate to the data folder and open the Yemen_Admin2.json file.



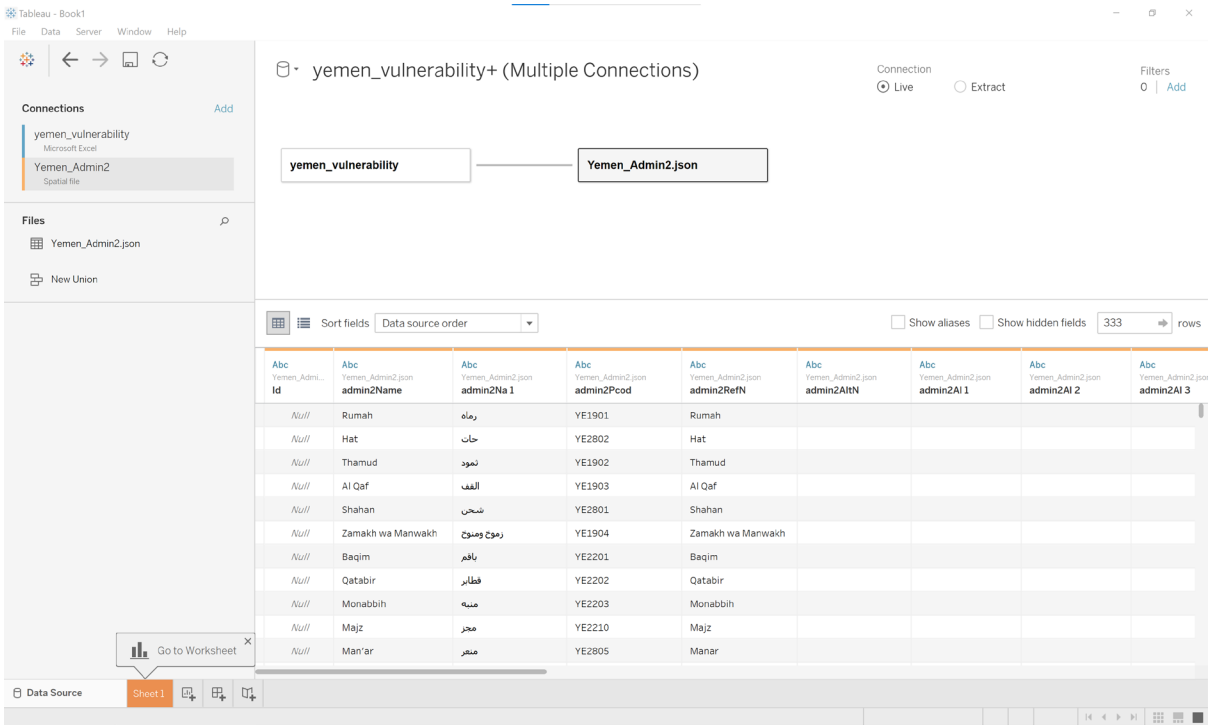
6. The new Spatial JSON file will be presented under the yemen_vulnerability Excel file.



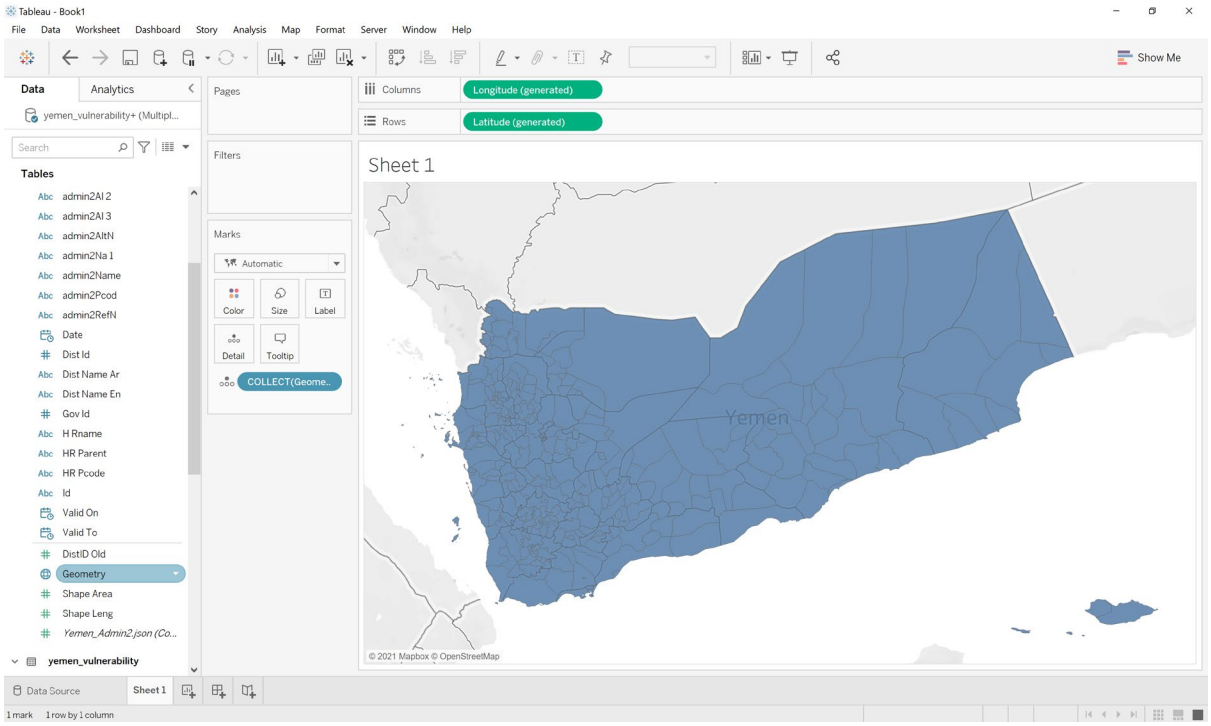
7. Drag the Yemen_Admin2.json file into the data window. An Edit Relationship window will open. As both datasets share the same PCODES attribution it is possible to join the data. This will allow us to analyse attributes from both datasets together. For yemen_vulnerability select District Pcode, for Yemen_Admin2.json select admin2Pcod.



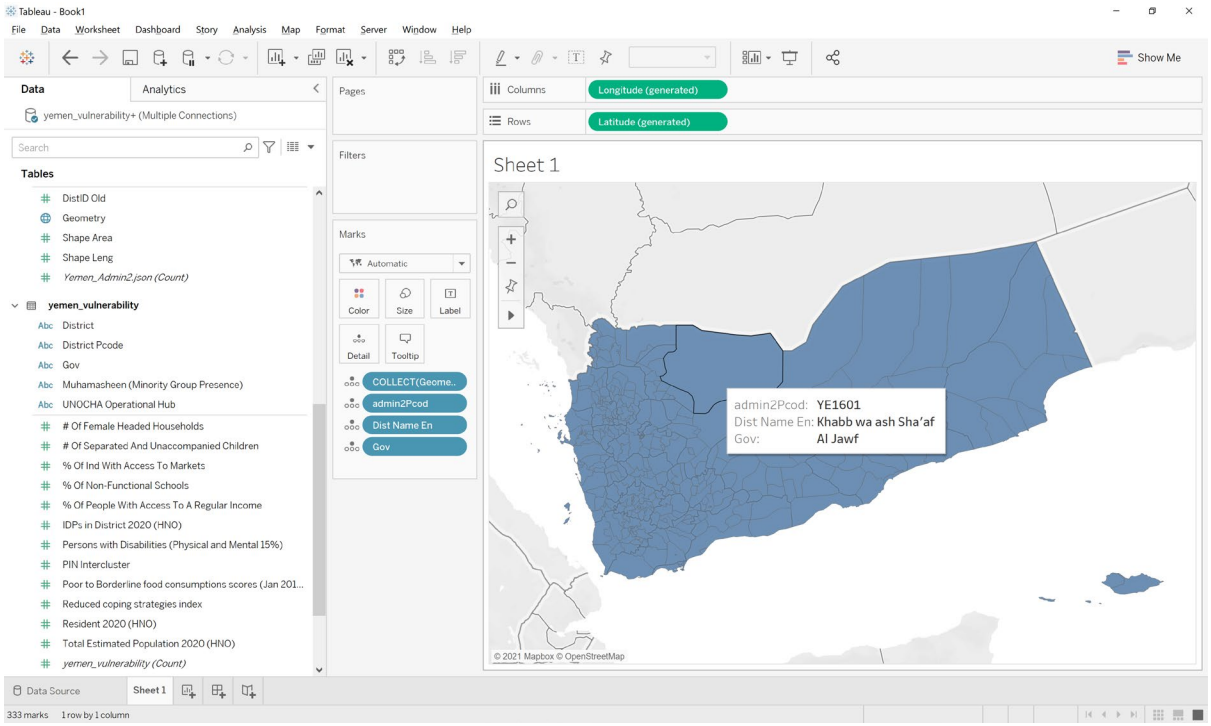
8. Once the datasets are joined, click the Sheet 1 tab at the bottom of the window.



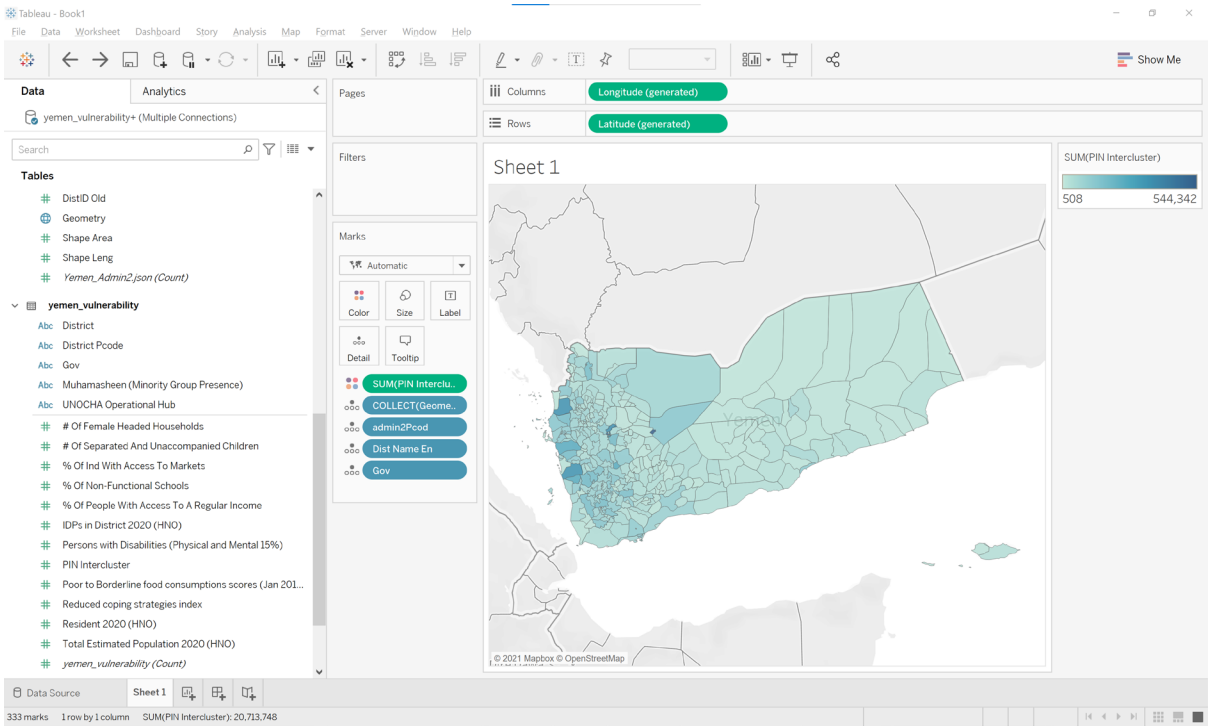
9. Once navigated to the Sheet 1 tab, locate the Geometry field and drag into the data window.



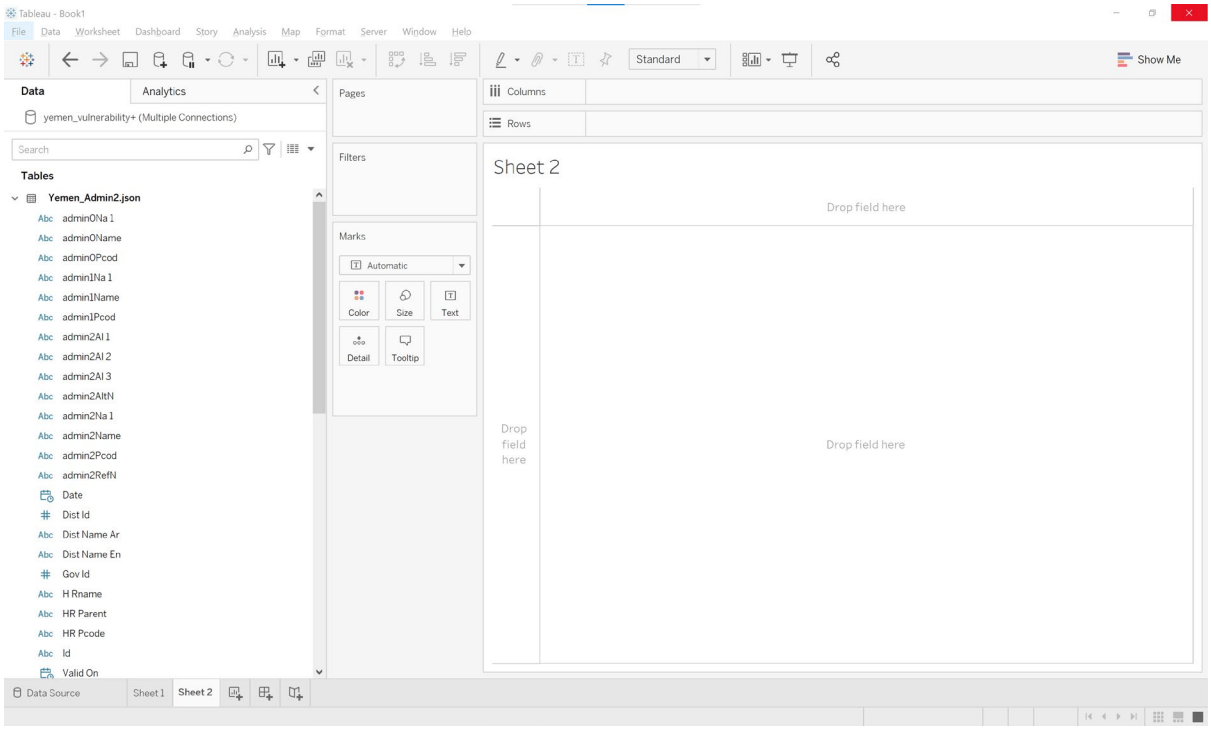
10. To be able to select individual districts, drag and drop the fields ‘admin2Pcod’, ‘Dist Name En’ and ‘Gov’ onto the Detail mark.



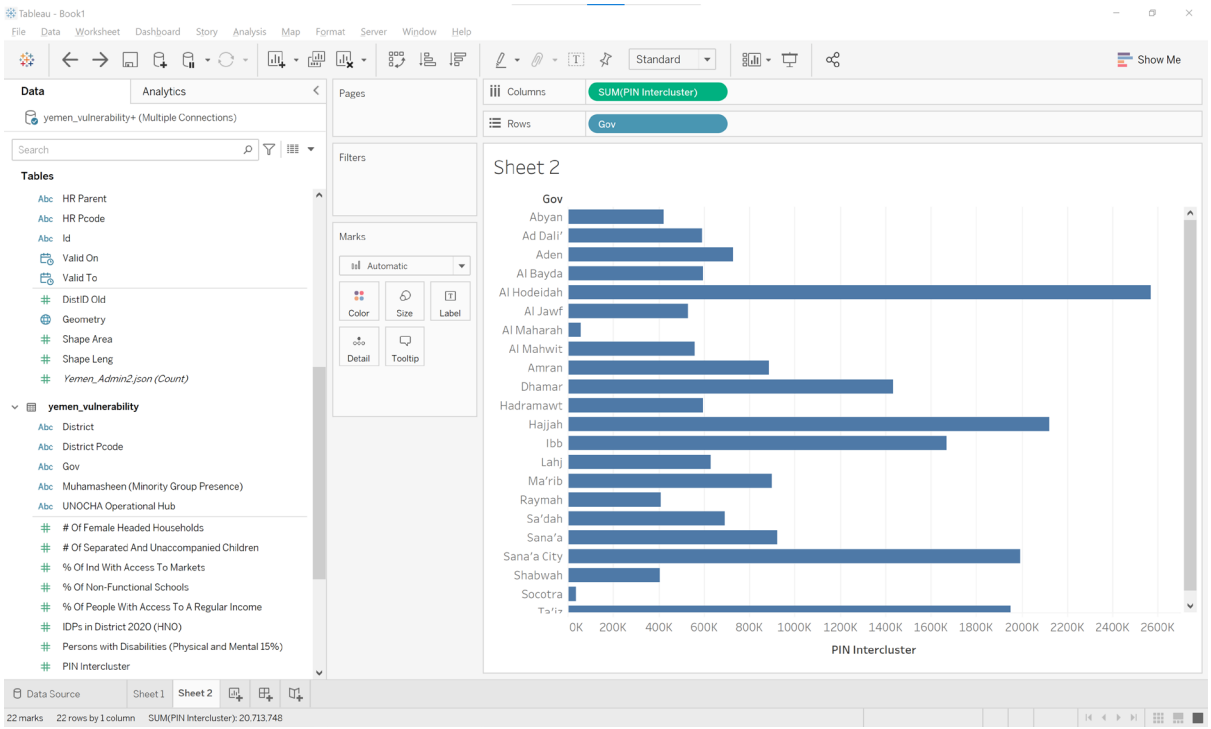
11. To visualise the total number of people in need at a district level, drag and drop the field ‘PIN Intercluster’ onto the Color mark.



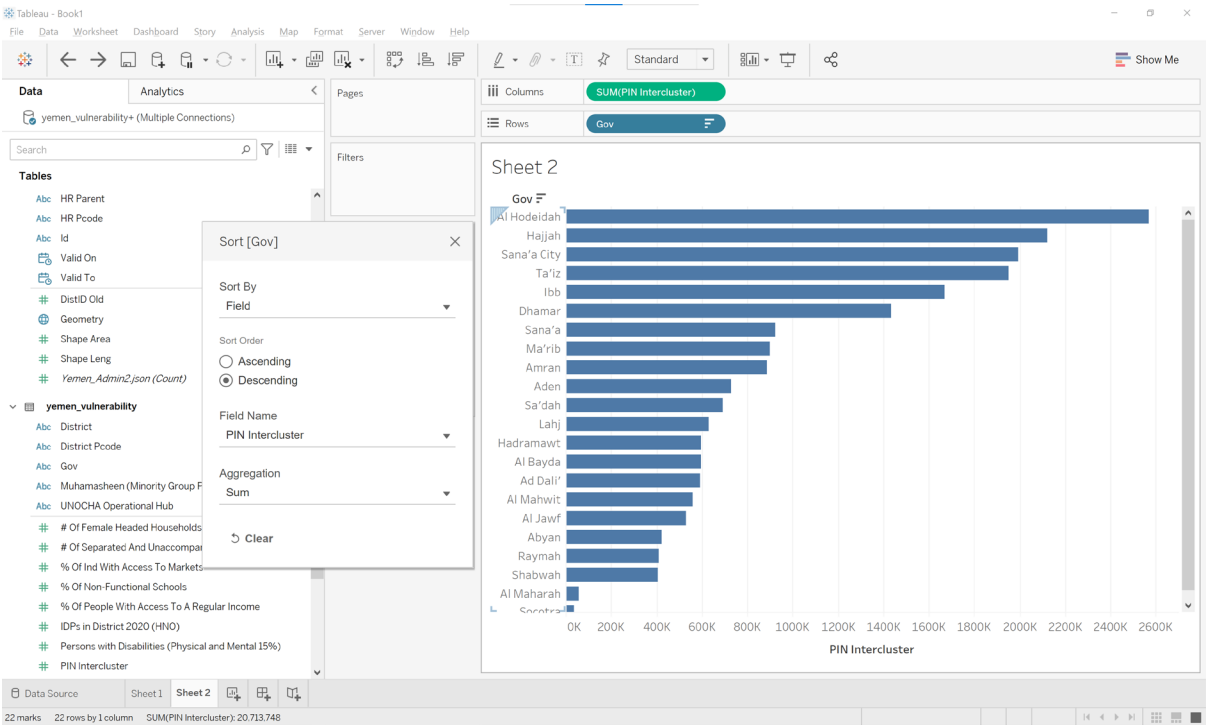
12. Next, we will make a graph of the number of people in need per governorate. At the bottom of the desktop window click the 'New Worksheet' icon  . This will open a new blank sheet.



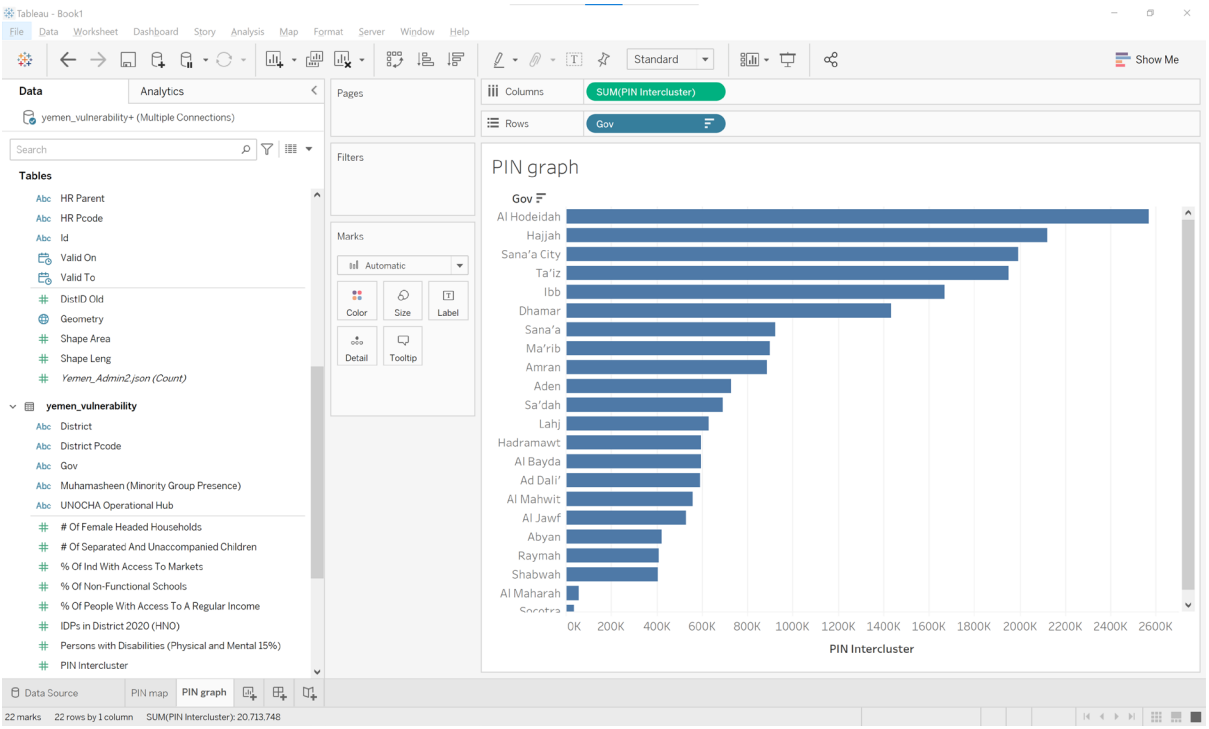
13. Drag and drop field 'PIN Intercluster' onto Columns. Drag field 'Gov' onto Rows.



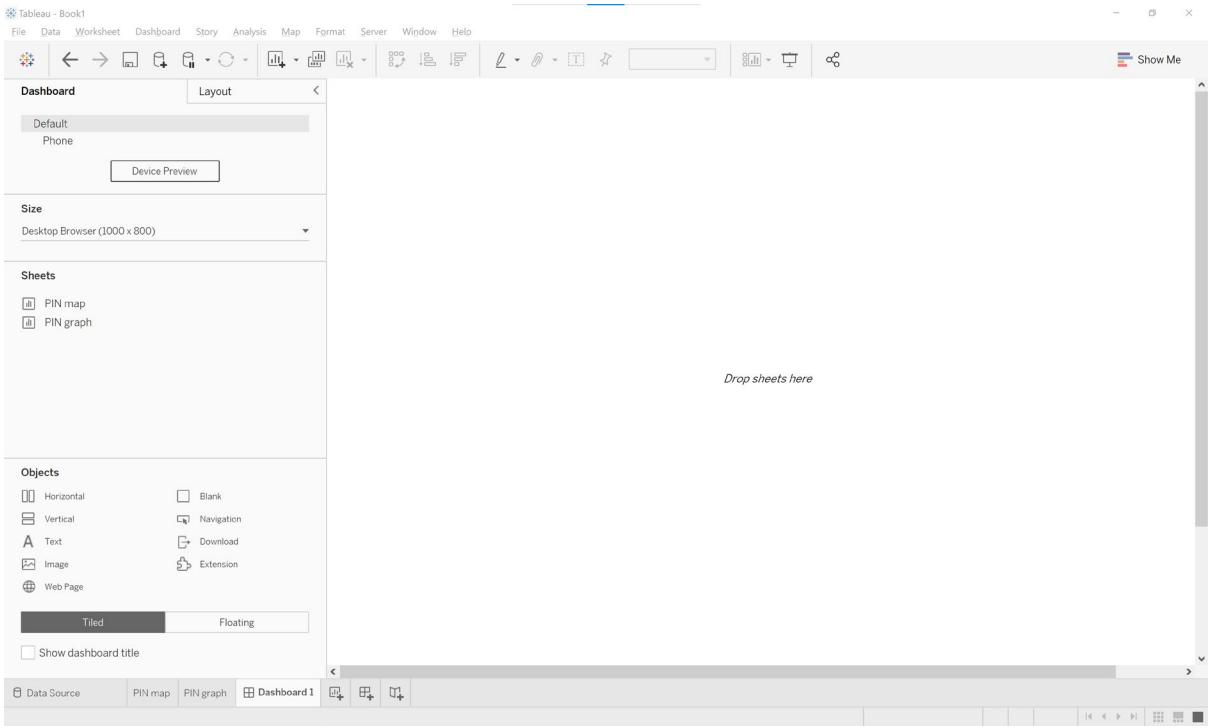
14. Right click on 'Gov' and select Sort. Select the following options.
    Sort By: Field
    Sort Order: Descending
    Field Name: PIN Intercluster
    Aggregation: Sum



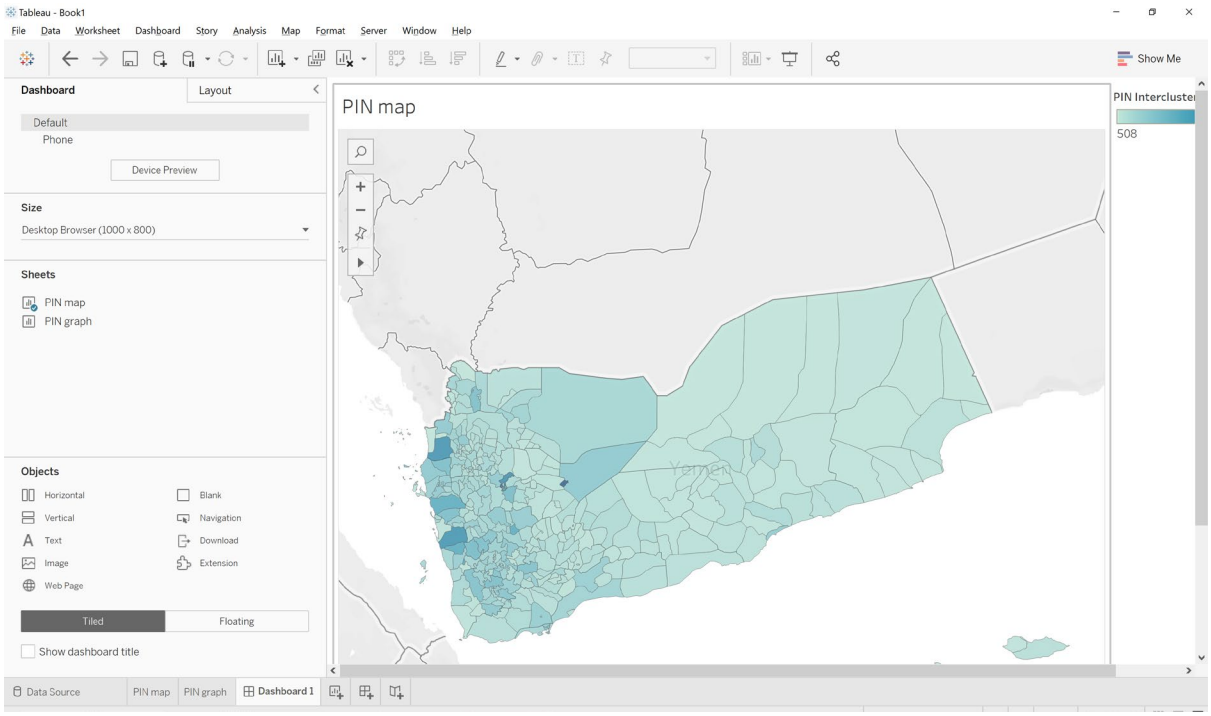
15. Close Sort window. On the Sheet bar at the bottom, double click Sheet 1 and rename to PIN map. Double click Sheet 2 and rename to PIN graph.



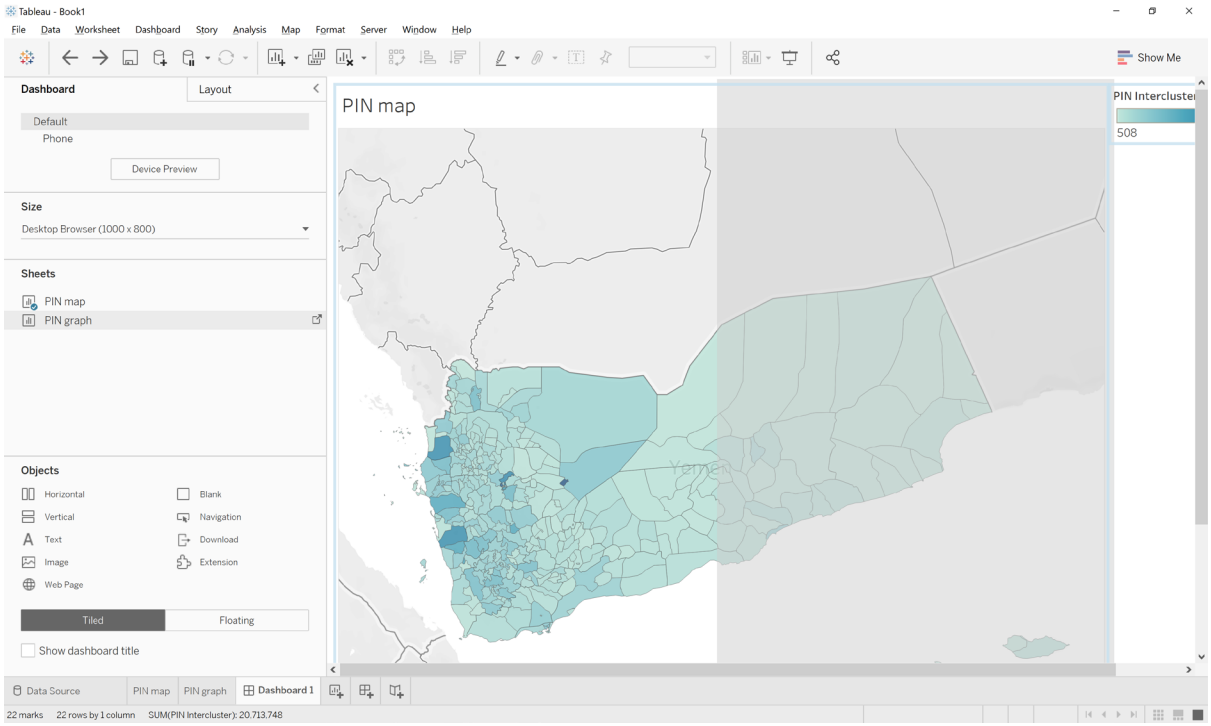
16. At the bottom of the window click the New Dashboard icon  . A new dashboard window will open.



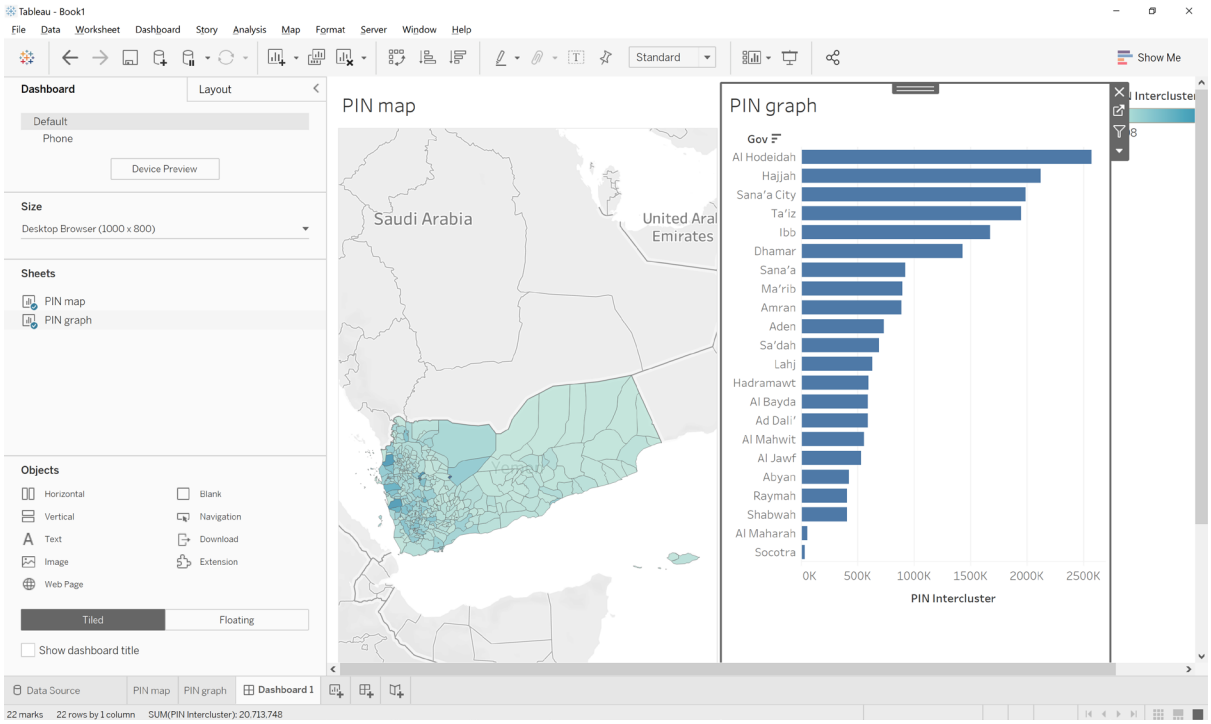
17. Under the Sheets window, drag 'PIN map' into the dashboard window.



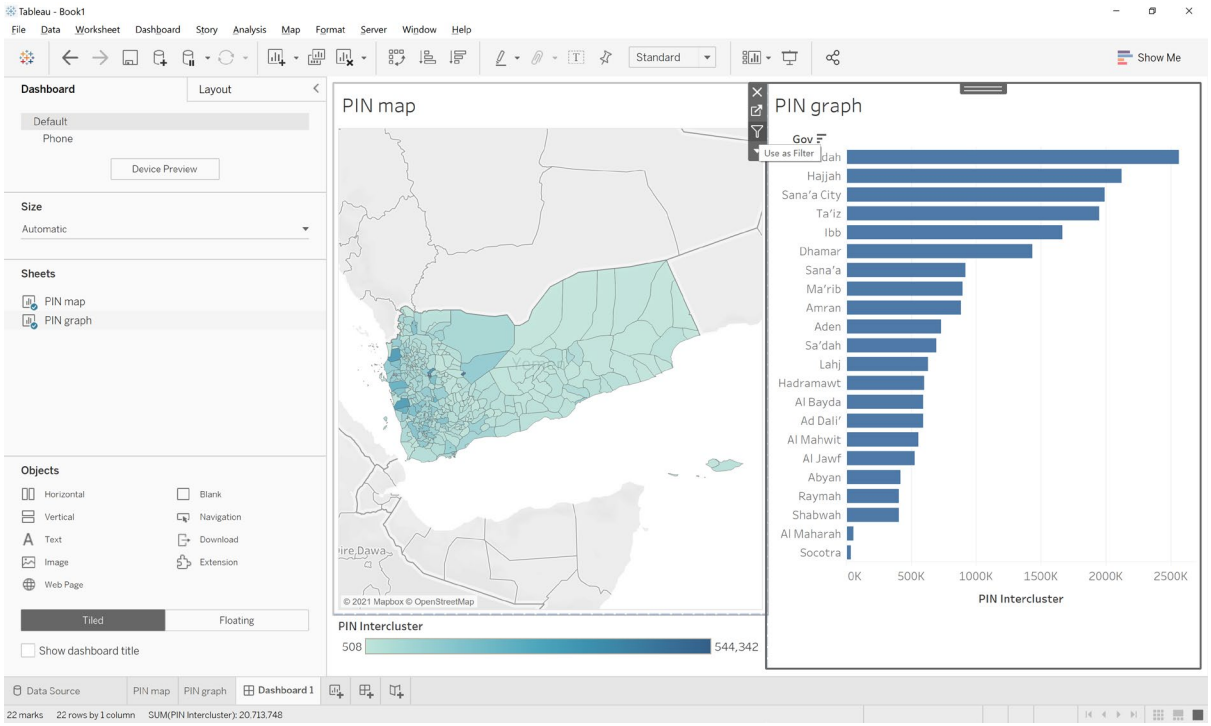
18. Under the sheets window, drag 'PIN graph' into the dashboard window.



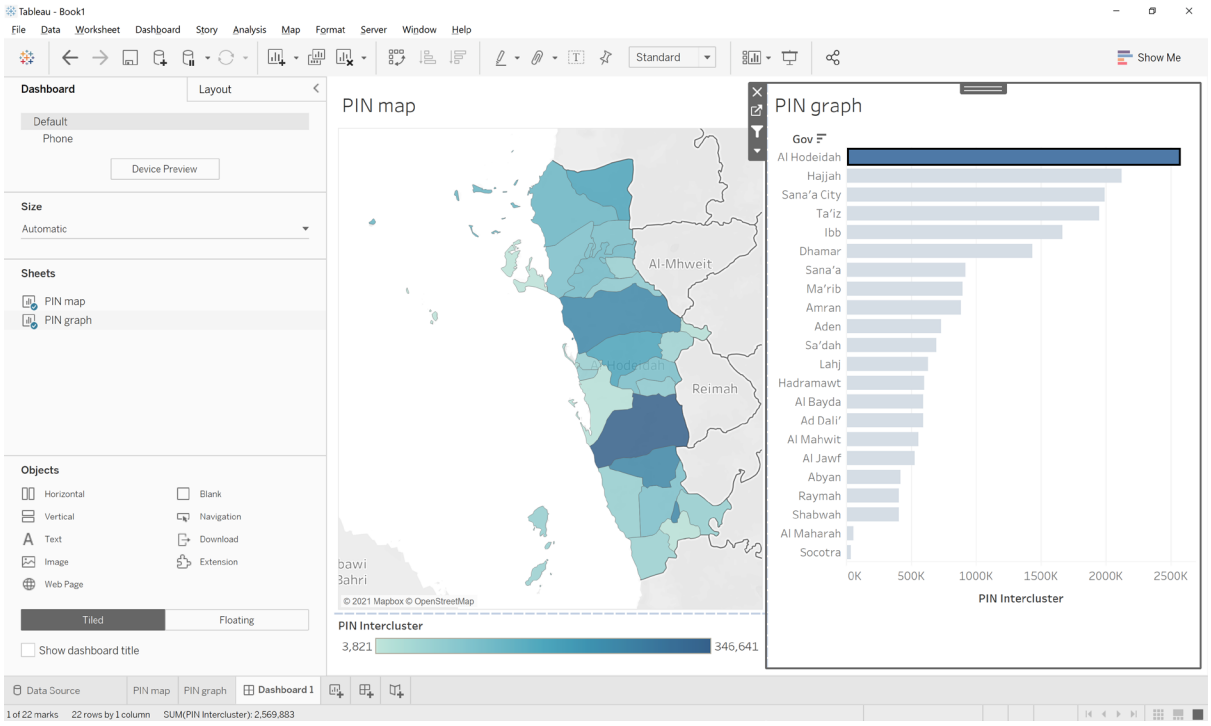
19. Both the graph and map will now be visible on the dashboard. Rearrange sheets if required.



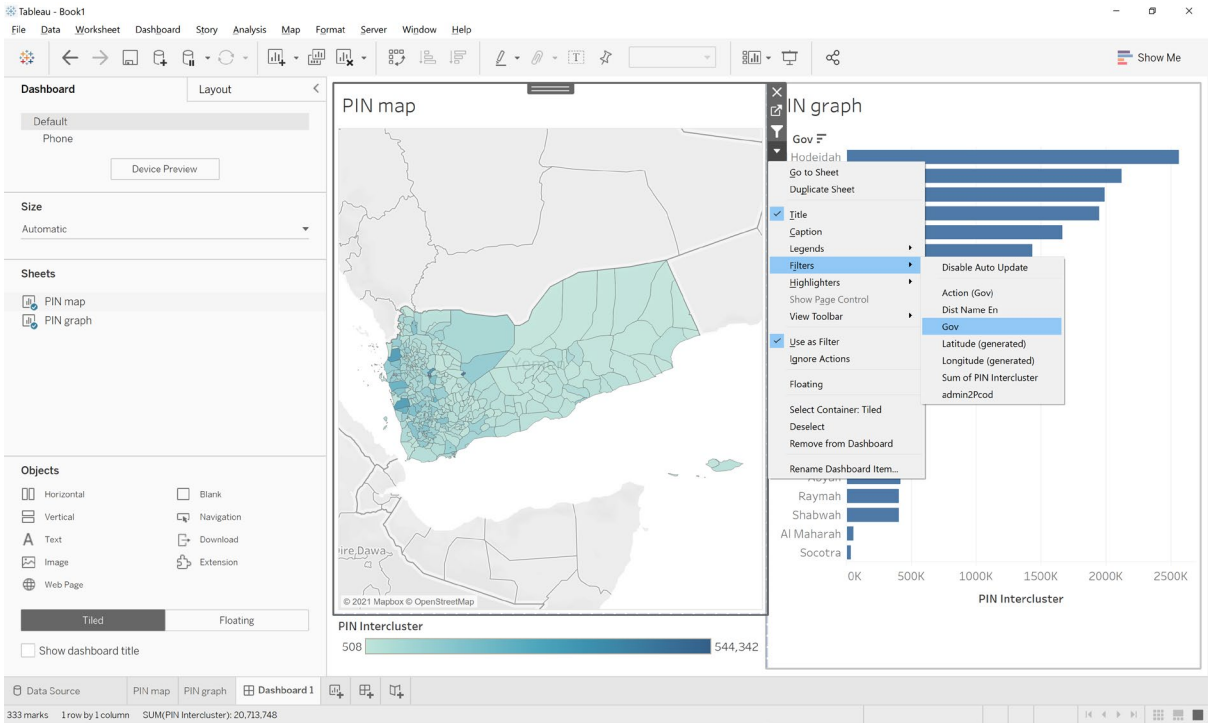
20. Click on the PIN graph and click on the 'Use as Filter' icon.



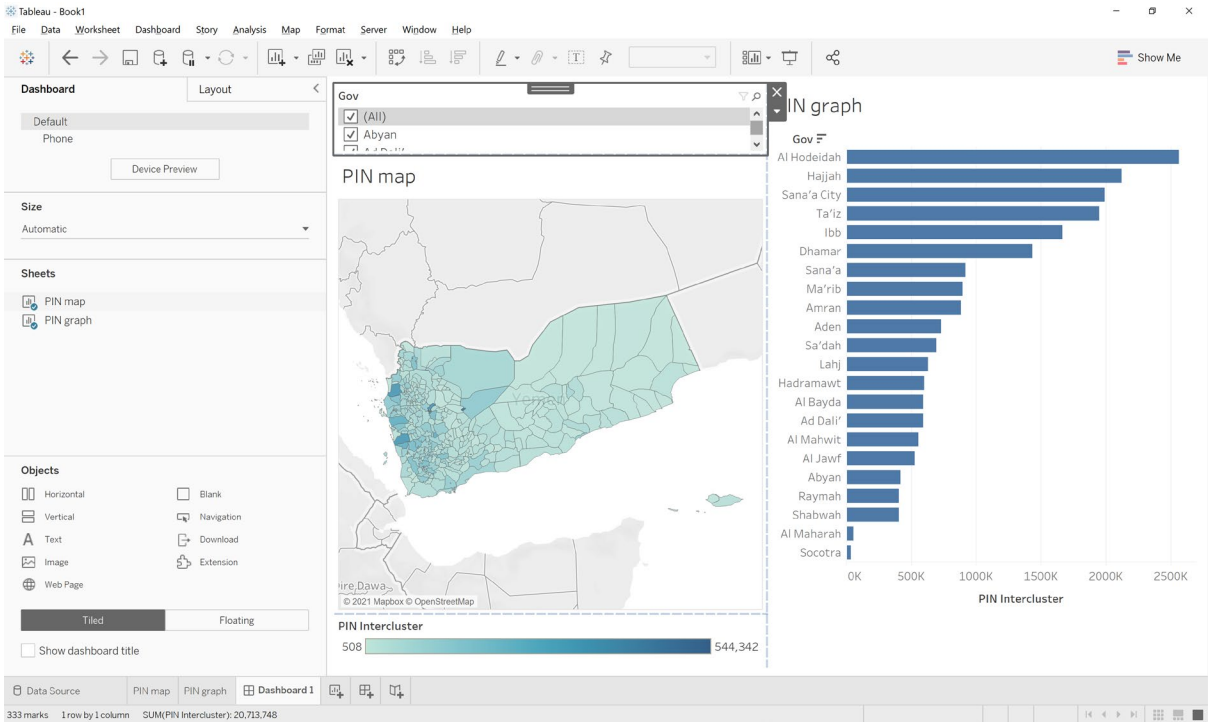
21. Once the filter is enabled, when a governorate is clicked on the graph, the map is also filtered.



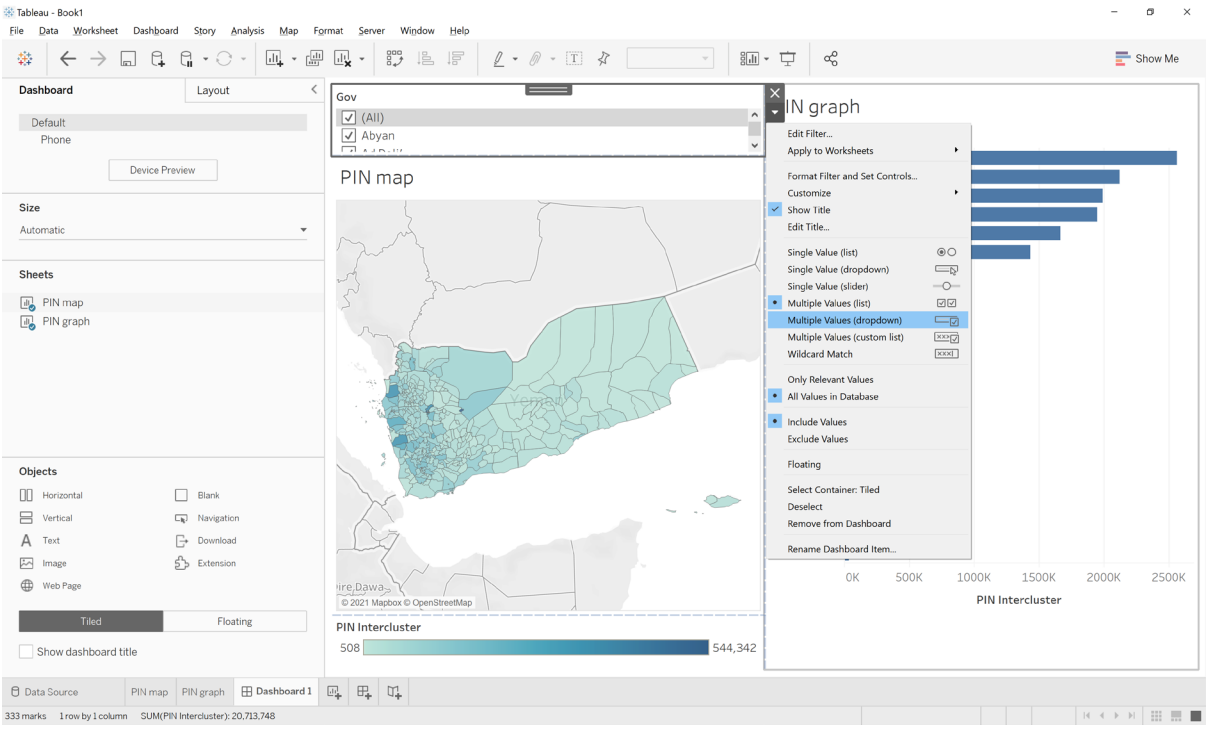
22. To create a drop down Governorate filter, click on the map and the down arrow button. Then click Filters -> Gov.



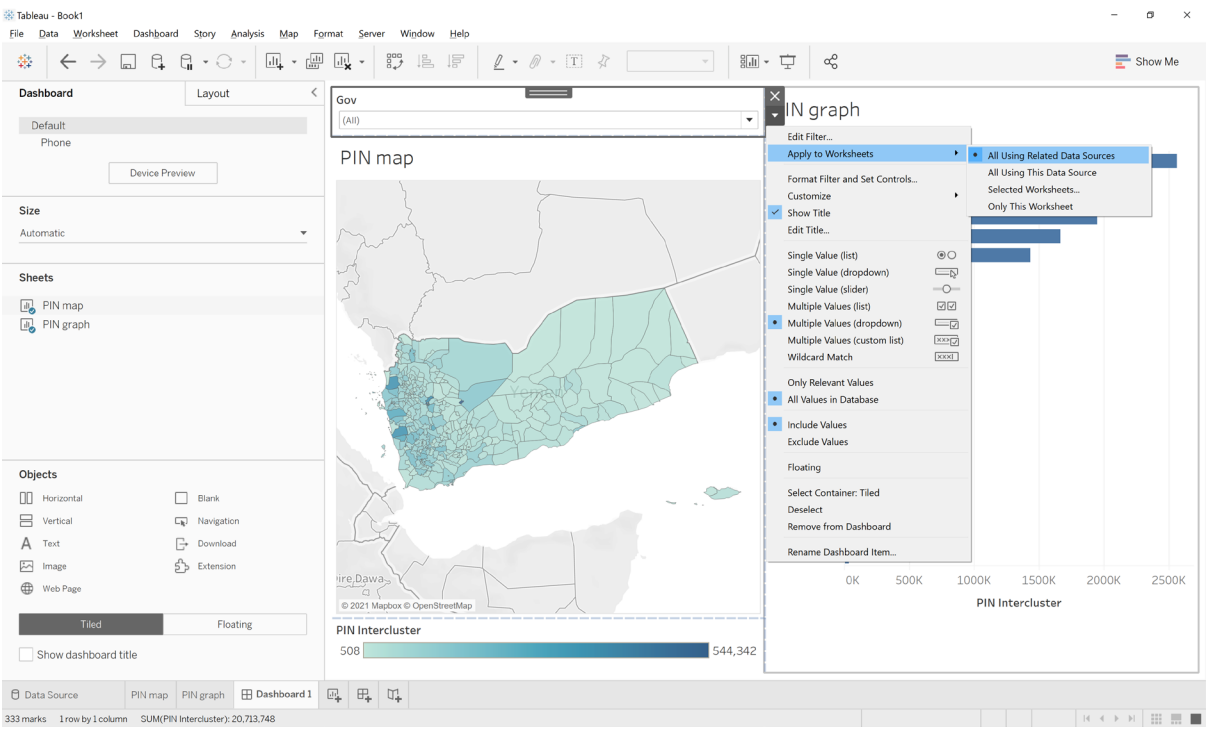
23. A new filter box will be created.



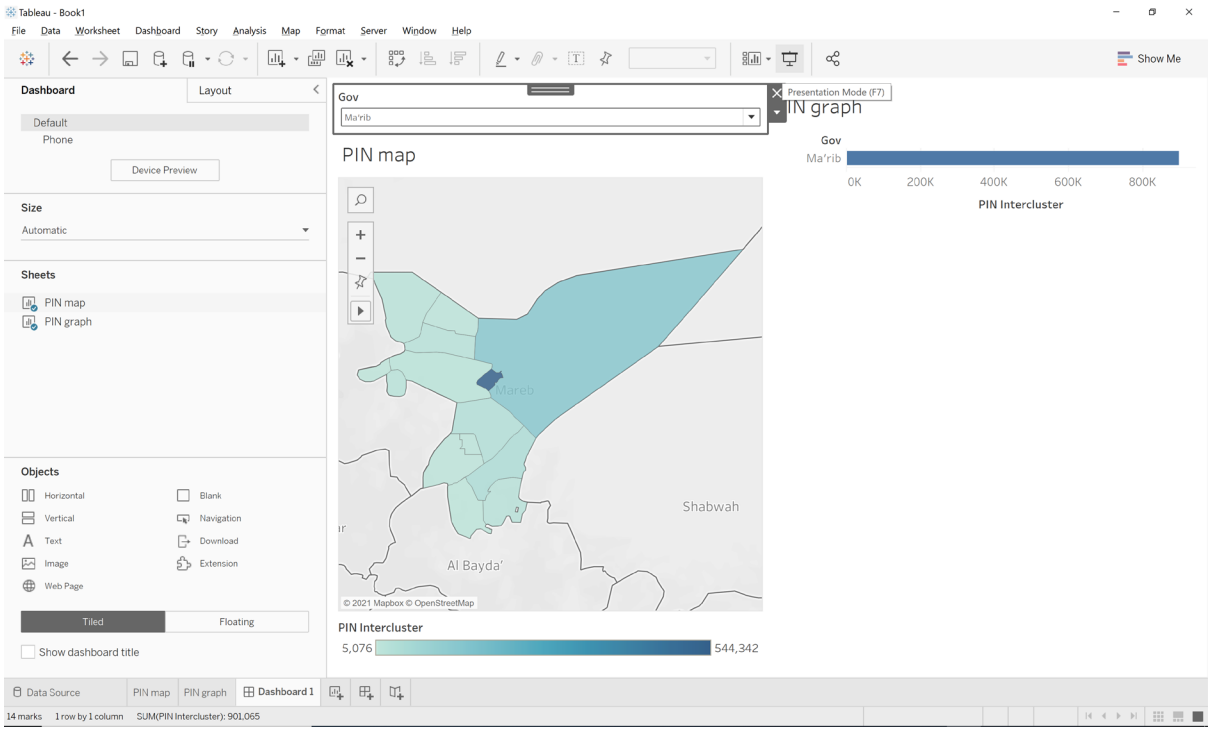
24. Select the drop down arrow on new filter box and select 'Multiple Values (dropdown)'. This will change the style of the filter.



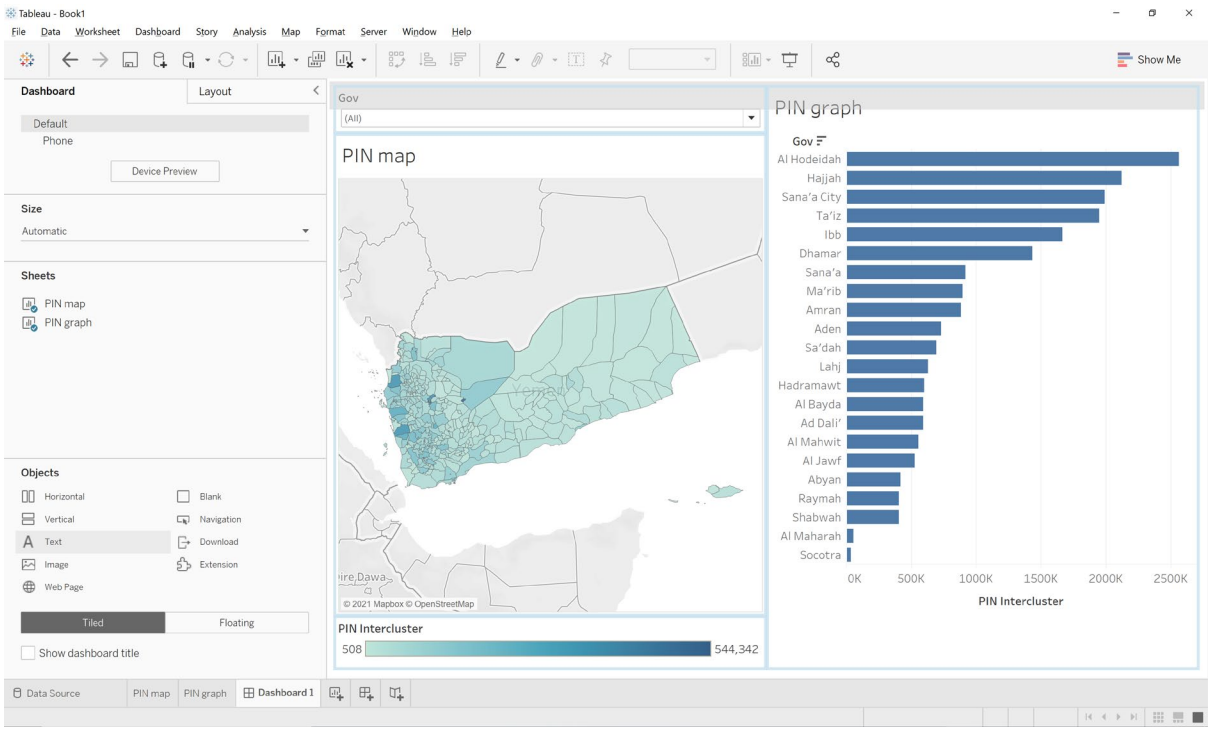
25. On the filter, select the drop down arrow and click 'Apply to Worksheets' -> 'All Using Related Data Sources'. This will ensure the filter will be related to the graph and map.



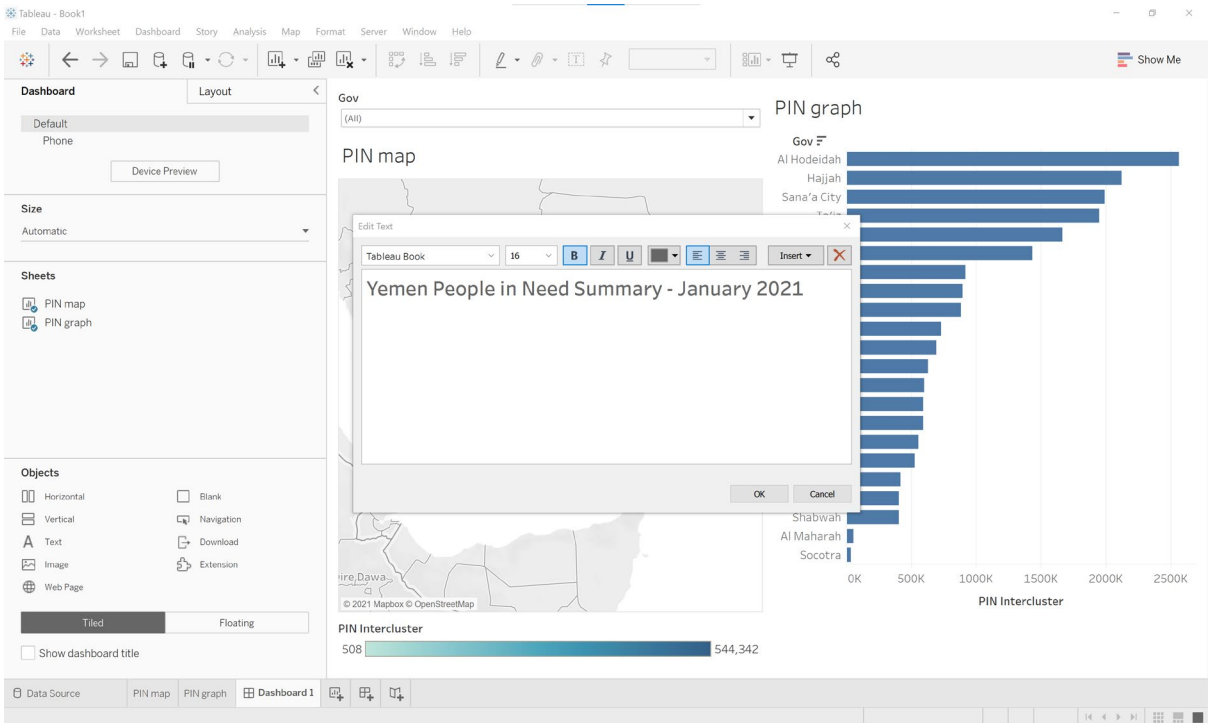
26. Selecting a governorate from the Gov filter will now filter both the map and the graph.



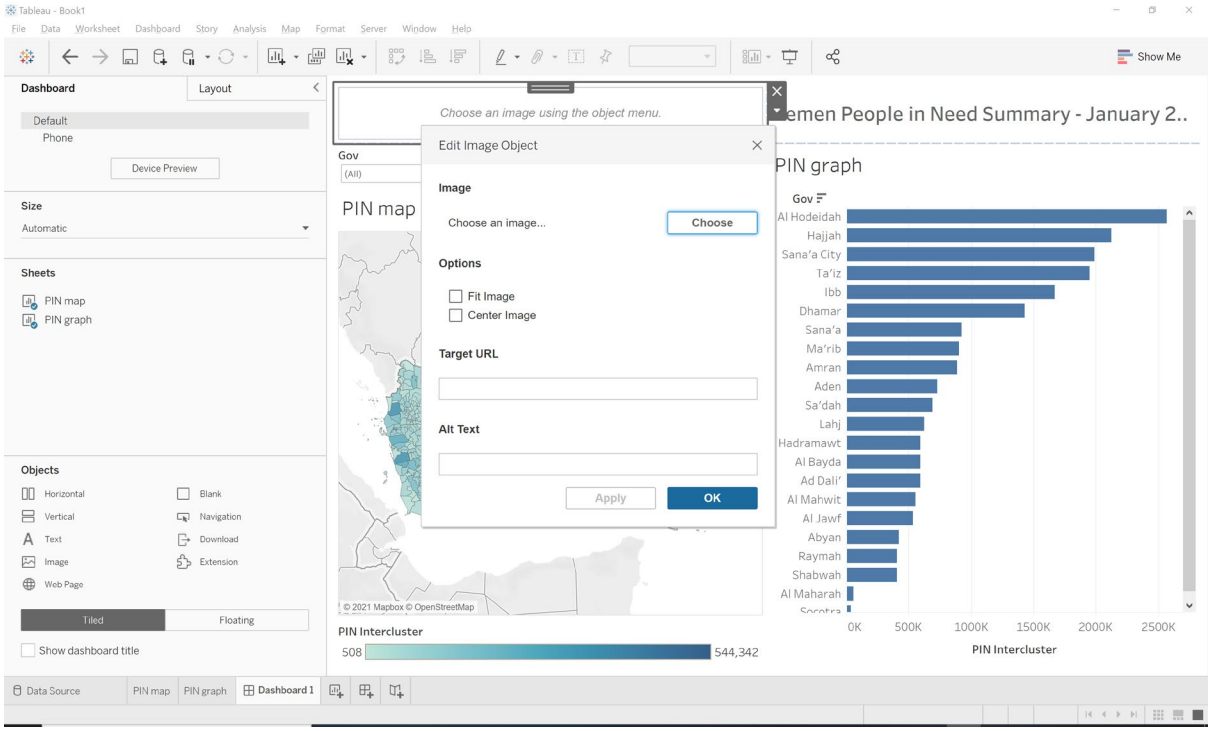
27. To add a title to the dashboard, drag and drop a Text object into the dashboard window.



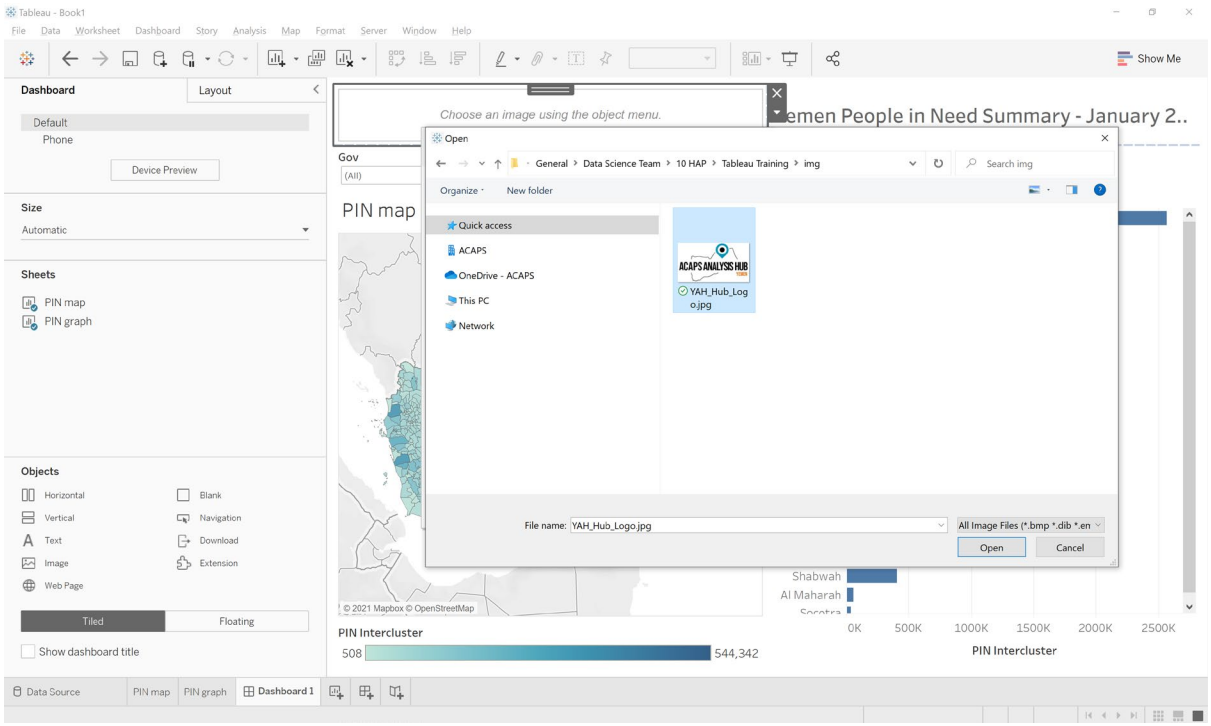
28. Add text title such as 'Yemen People in Need Summary – January 2021'.



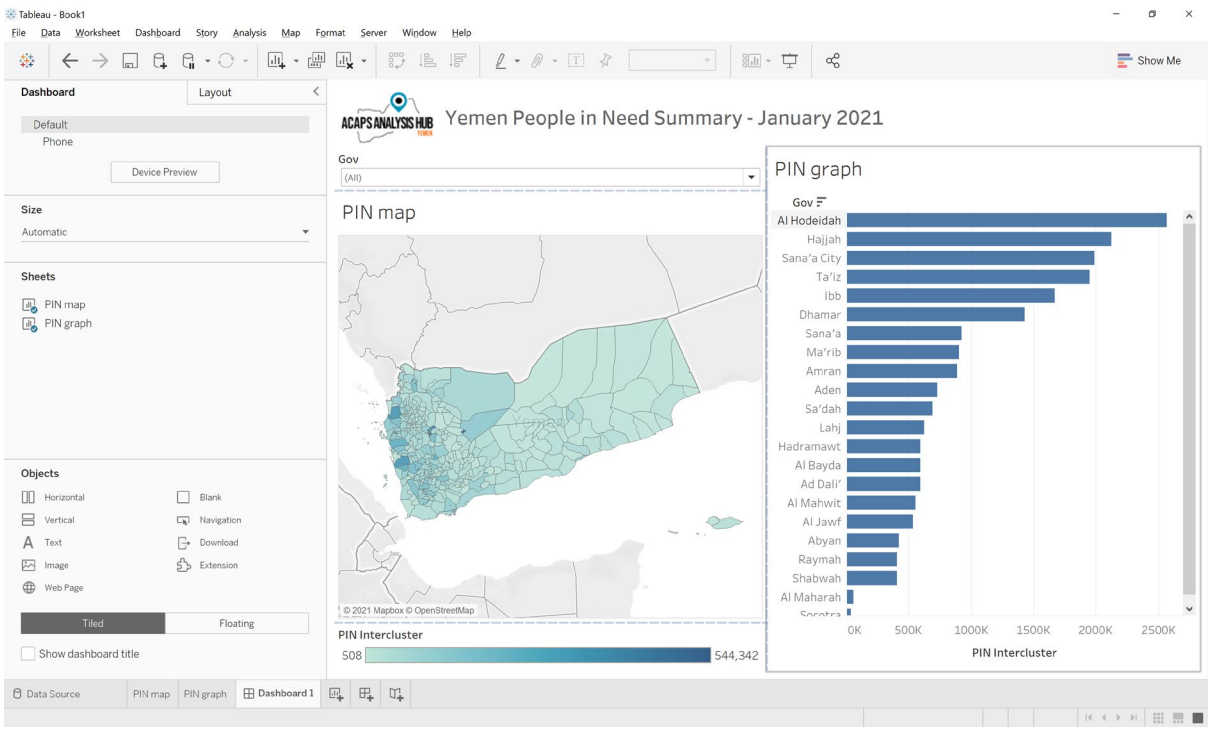
29. To add a logo to the dashboard, drag and drop an Image object into the dashboard window.



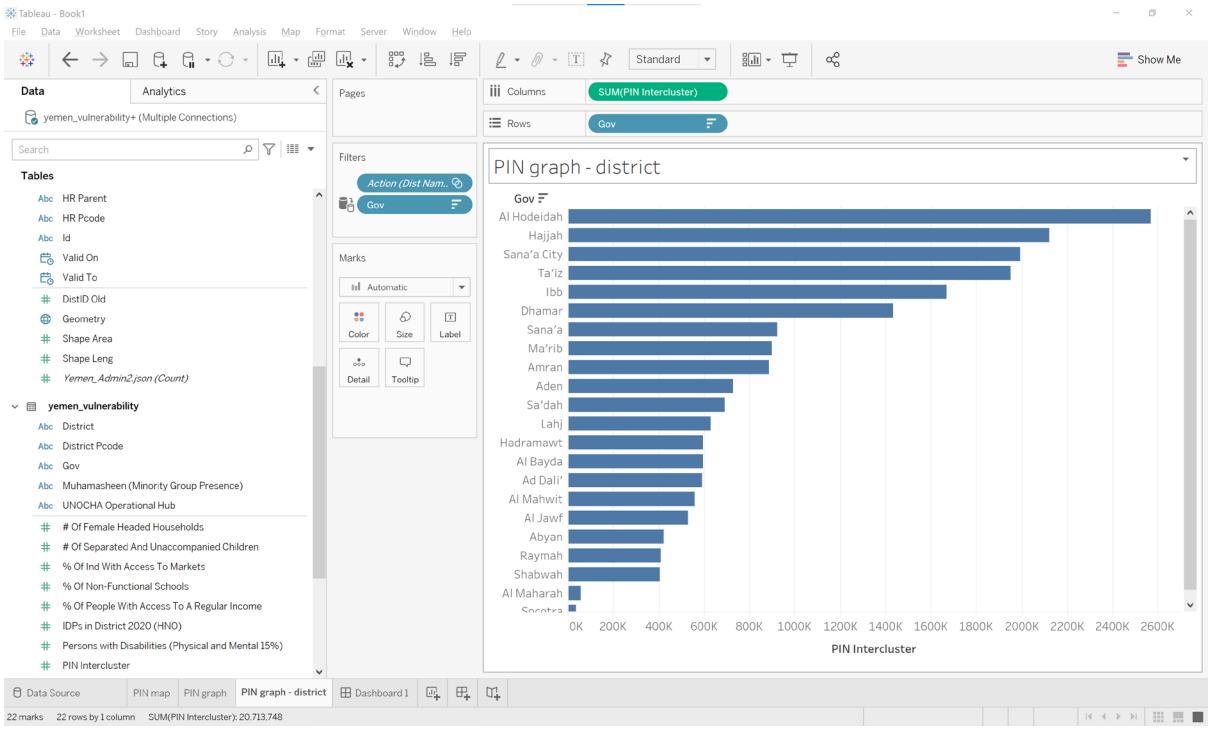
30. Click choose and navigate to your logo directory (under Tableau Training -> img).



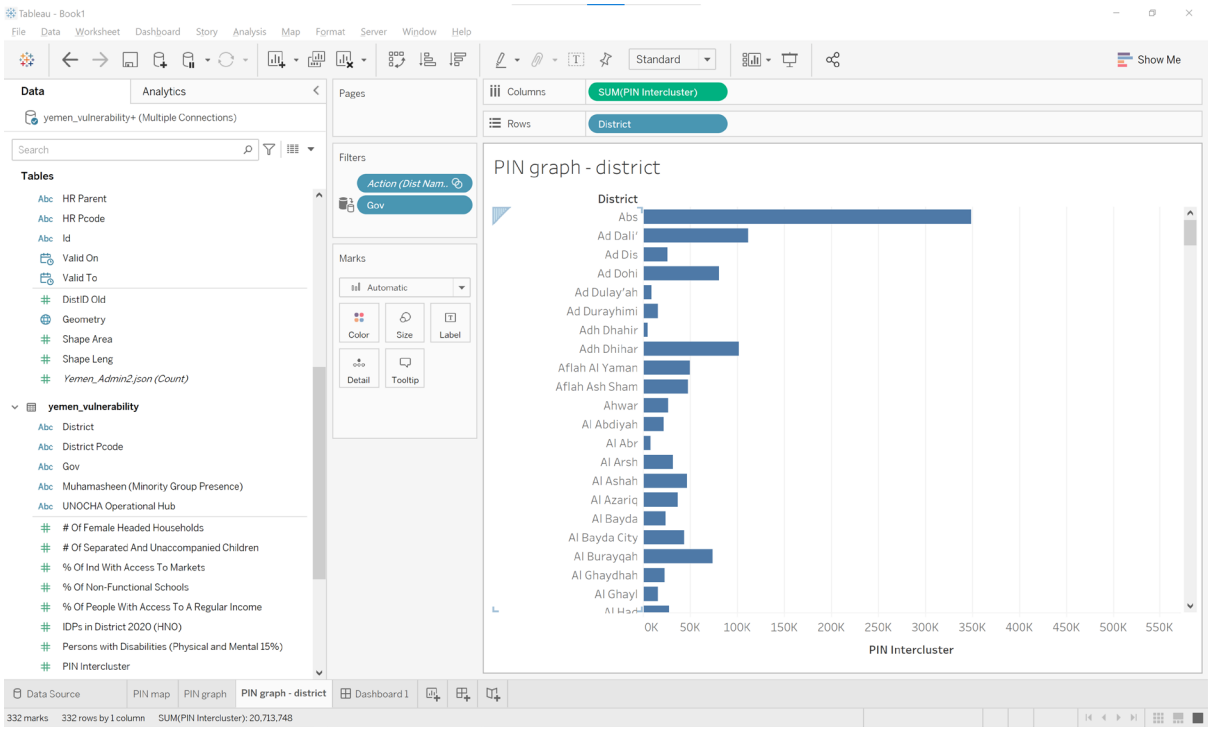
31. The dashboard now shows the highest number of people in need by governorate. This can be easily filtered via the map or the graph. However, it does not show us PIN at a district level. To show at a district level instead we will need to create a new graph.



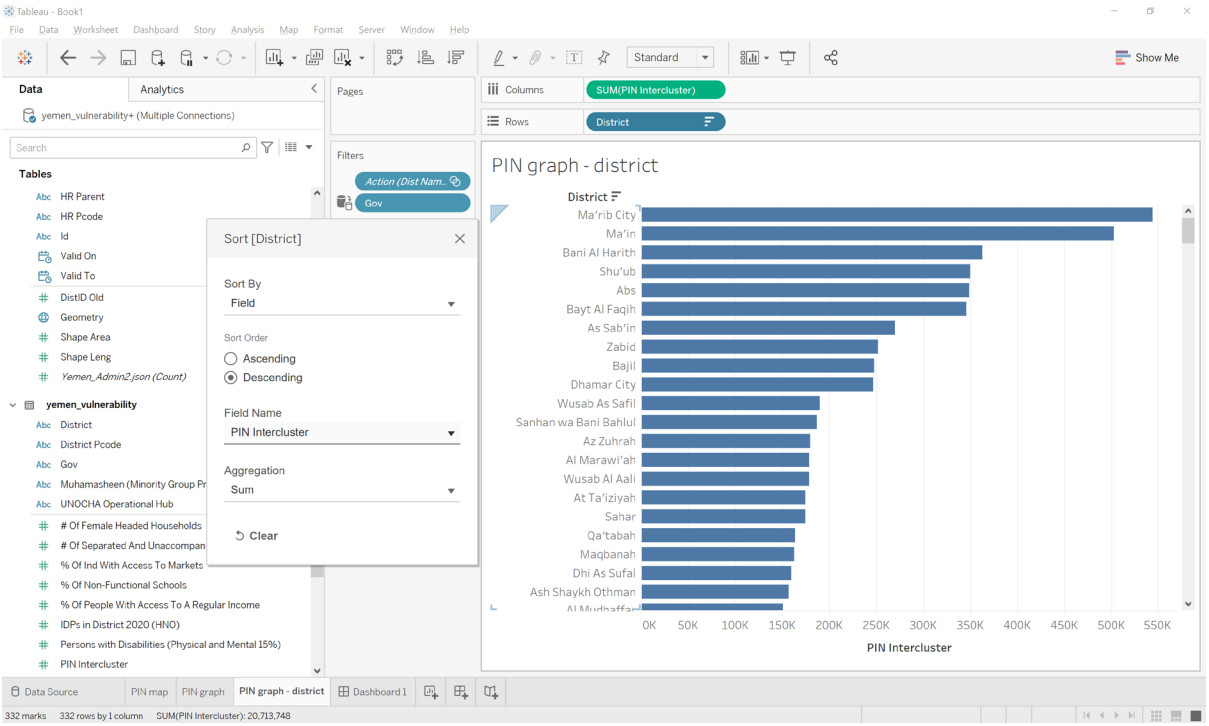
32. Right click on ‘PIN graph’ and click duplicate. A new version of the graph will be created. Double click on the tab and rename to ‘PIN graph – district’.



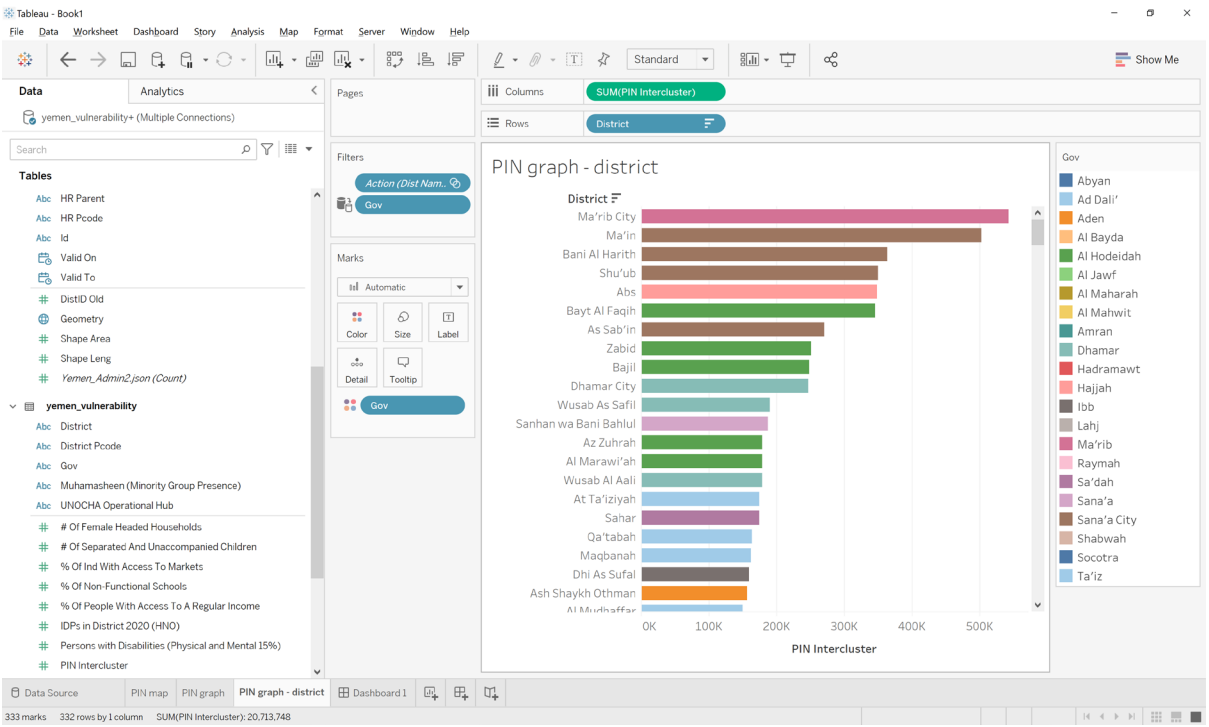
33. Replace ‘Gov’ through dragging the field ‘District’ onto rows.



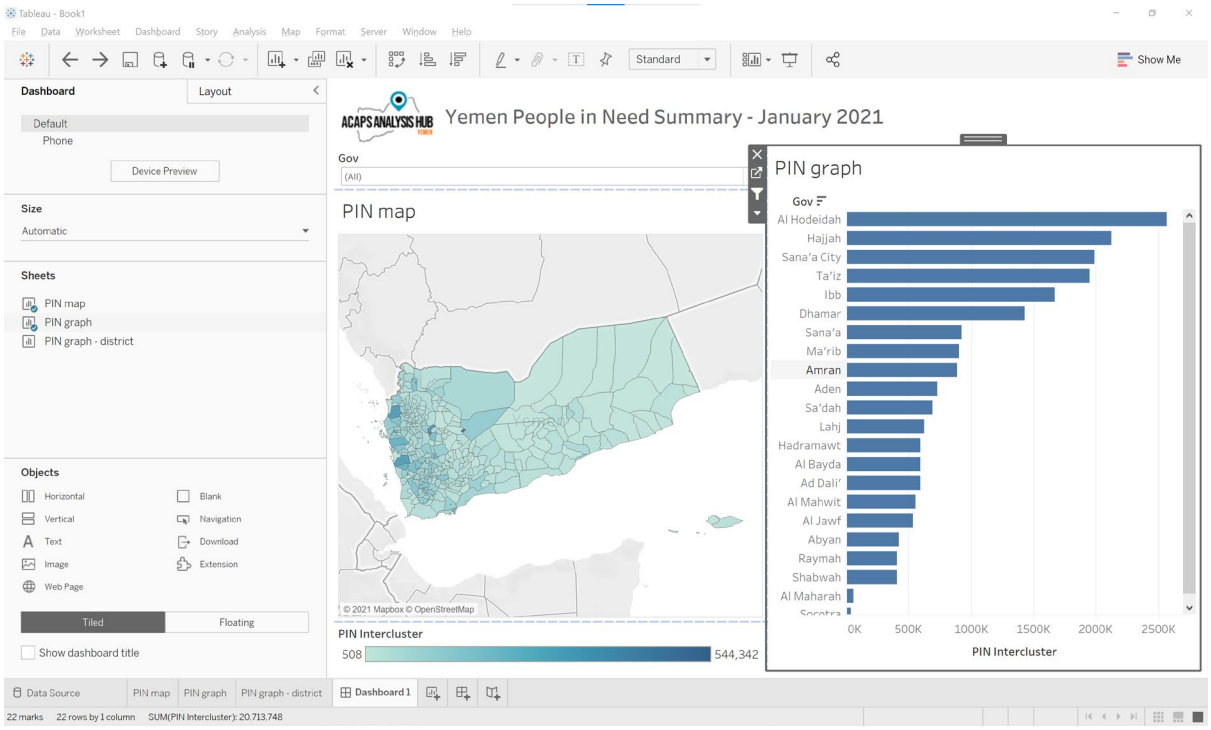
34. Right click on 'District and select Sort. Select the following options.
    Sort By: Field
    Sort Order: Descending
    Field Name: PIN Intercluster
    Aggregation: Sum



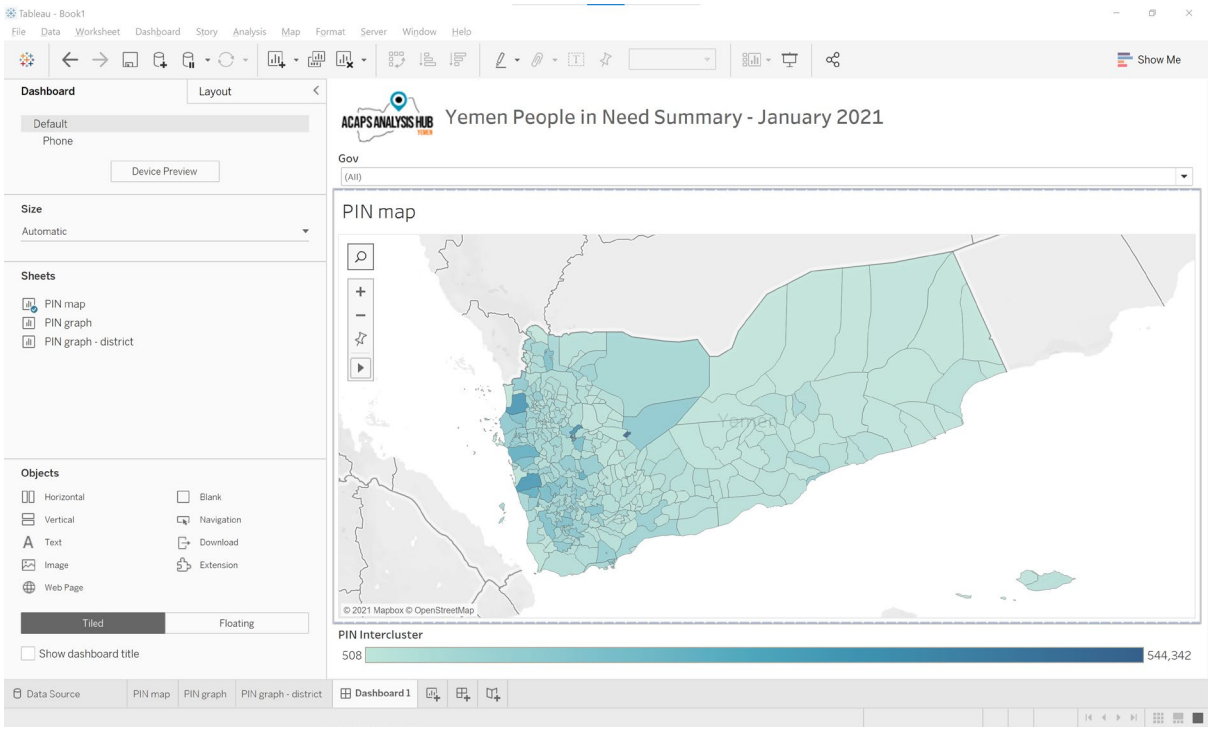
35. Drag the field 'Gov' onto the Color mark. This will colour each district by governorate.



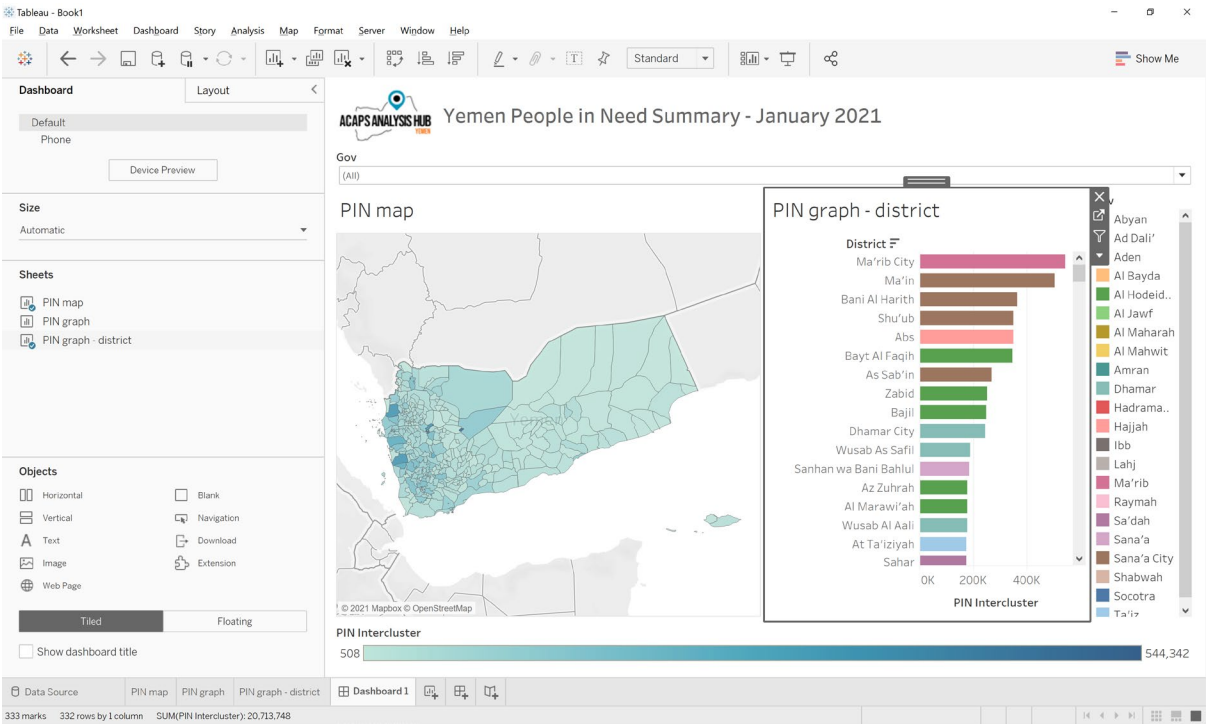
36. Return to the dashboard tab, click on the PIN graph and click the cross to remote the Governorate graph from the dashboard.



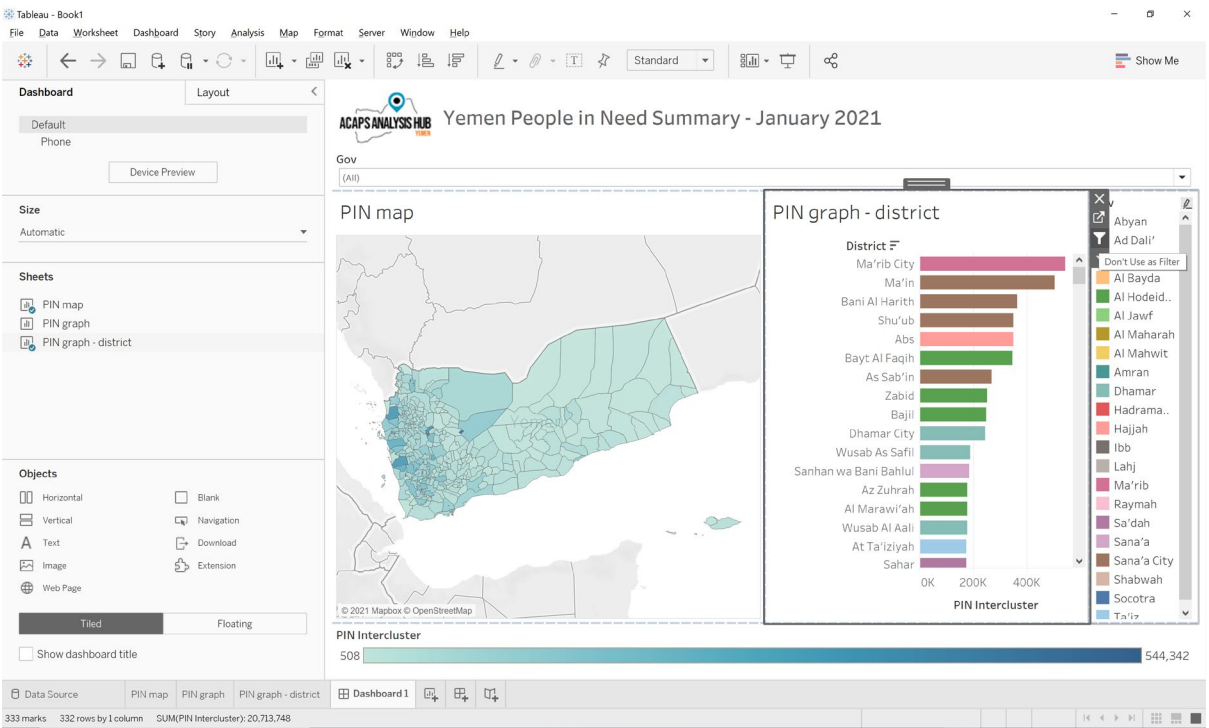
37. Drag and drop the new 'PIN graph – district' onto the dashboard..



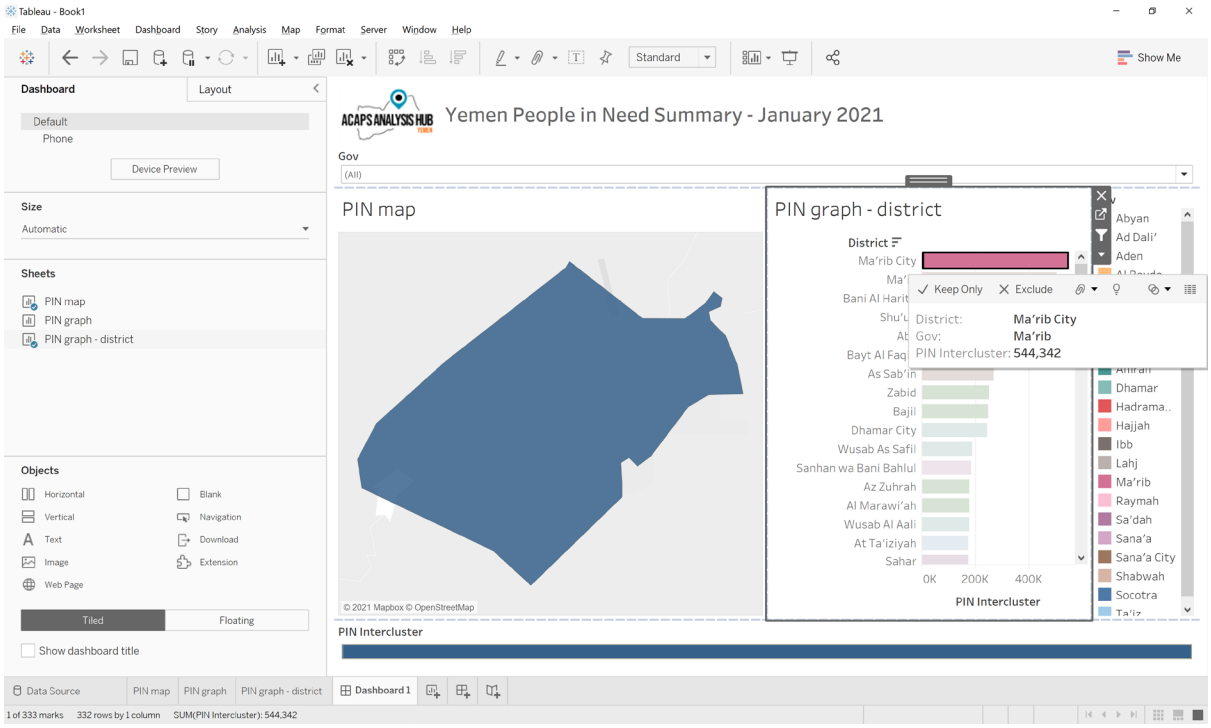
38. Rearrange the dashboard as necessary. The dashboard should now contain a map, a graph and a governorate legend.



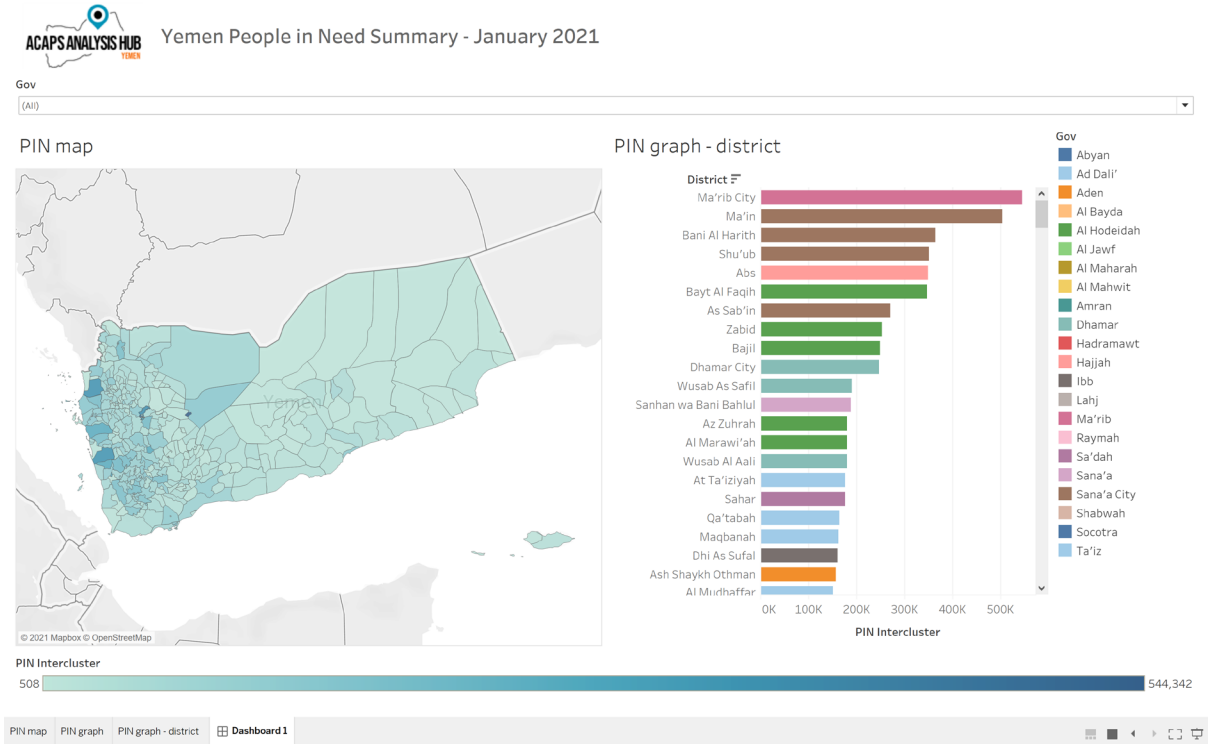
39. Click on the 'PIN graph – district' and click on the filter icon.



40. It is now possible to click on the graph to filter by district.



41. The dashboard is now complete. It is now possible to analyse and filter PIN numbers by district using both the map and the graph.



42. To save your dashboard, click File -> Save As and save in a suitable location. Tableau files are saved as .twb files.

Additional exercises

Now you have completed your first operational dashboard, attempt to answer the following questions by adding additional functionality and analysis to your dashboard:

1. Using fields 'IDPs in District 2020 (HNO)' and 'Total Estimated Population 2020 (HNO)', identify which district has the highest number of IDPs as a percentage of population? Note: Total Estimated Population = IDPs in District + Resident. Any Districts where this does not add up suggests incorrect data and should be flagged or removed from the analysis
2. Using fields '% of people with access to a regular income' and 'Total Estimated Population 2020 (HNO)', calculate the 5 top districts which have the highest estimated population without access to a regular income. What might be the limitations with this methodology?
3. Using the 'Muhamasheen (Minority Group Presence)' field, create a table, map and filter visualising which districts have Muhamasheen present and which districts do not

Further training resources

Further training resources can be found at https://www.tableau.com/learn/training/20211.

Appendix 1 – Data review

Field	Description
District Pcode	PCODES are unique admin area codes use to define a geographical area such as an governorate or district boundary. Using PCODES helps to ensure consistency between datasets as using a unique code mitigates against issues of different place names spellings and ensures datasets can be joined easily. PCODES are assigned by UNOCHA and are part of the Common Operational Datasets.
UNOCHA Operational Hub	Name of the UNOCHA Operational Hub used to plan OCHA humanitarian programmes.
Gov	Governorate name.
District	District name.
IDPs in District 2020 (HNO)	Number of internally displaced people (IDPs) within district as taken from the 2020 Humanitarian Needs Overview.
Resident 2020 (HNO)	Number of non-IDP residents within district as taken from the 2020 Humanitarian Needs Overview.
Total Estimated Population 2020 (HNO)	Total number of IDPs and residents within district as taken from the 2020 Humanitarian Needs Overview.
Persons with Disabilities (Physical and Mental 15%)	Approimation of the number of persons with disabilities by district.
# of separated and unaccompanied children	Number of separated and unaccompanied children by district.
# of female headed households	Number of female headed households by district.
Muhamasheen (Minority Group Presence)	Whether the district contains vulnerable Muhamasheen communities.
% of non-functional schools	Percentage of non-functional schools within the district.
% of people with access to a regular income	Percentage of people with access to a regular income.
Reduced coping strategies index	The Reduced Coping Strategies Index (RCSI) is a proxy indicator of household food insecurity. The higher the rCSI score, the worse the food insecurity as people are deemed to be using more negative coping strategies.
Poor to Borderline food consumptions scores (Jan 2019)	Number of people deemed to be have poor or borderline food consumption as of January 2019 by district.
% of ind with access to markets	Percentage of individuals with access to market by district.
PIN Intercluster	Number of intercluster people in need by district.