Lab-3 [Exploratory Data Analysis]

Out date: 22-Jun-2022
Due date: 25-Jun-2022, 11:59pm

Submission

1. Prepare your solutions in Tableau and save the workbook (e.g., Lab-3_LastName.twbx) **[50 points]**
2. Complete the tables given below and save the file (e.g., Lab-3_LastName.docx). **[50 points]**
3. Upload the files to the Canvas.

Objective: To carry out EDA on the given dataset, create effective visualizations and gain insights.

You will be using Tableau Desktop for this lab. Please download a trial version using the link in Canvas Information Page.

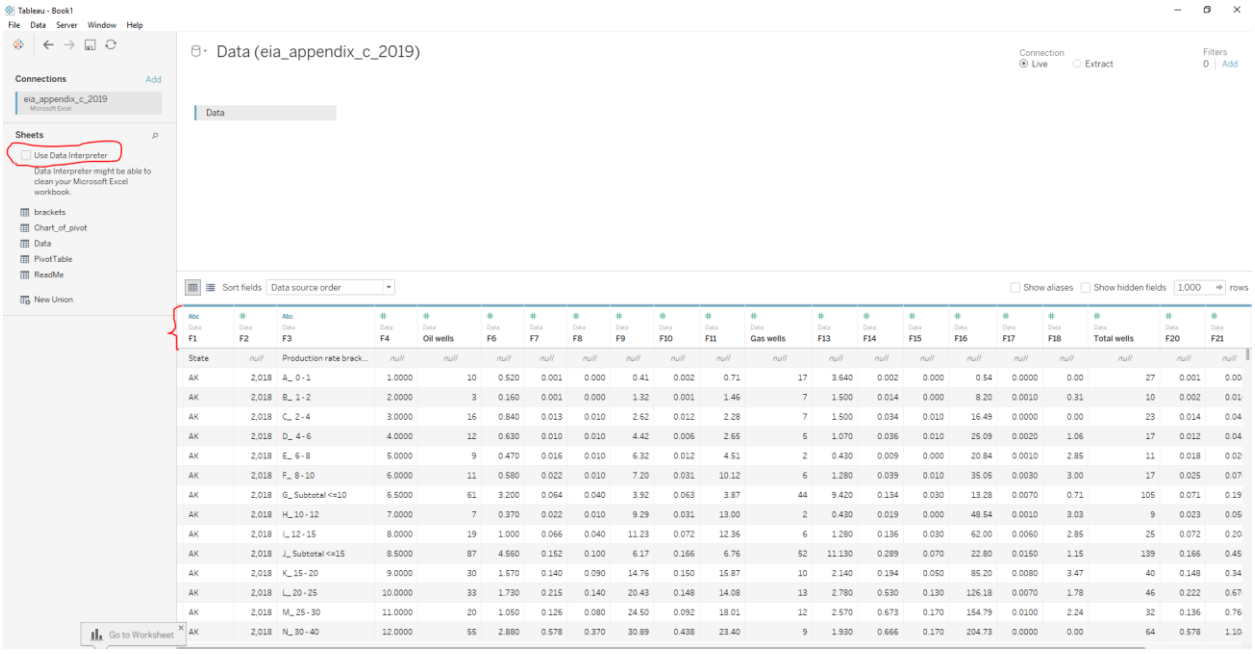
Problem 1/6: Locate and download the dataset [5 points]
(**Reference:** The data is used from U.S. Oil and Natural Gas Wells by Production Rate- https://www.eia.gov/petroleum/wells/.)

1. Read the first 3 paragraphs from the Introduction in the above link.

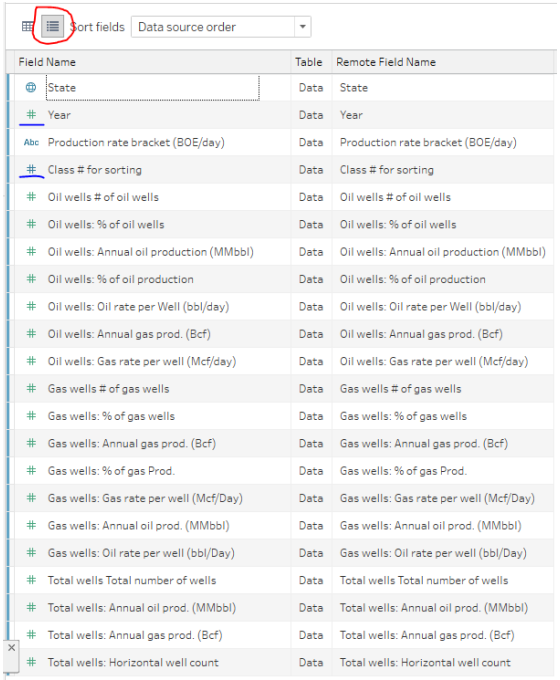
2. Download the Excel dataset **Appendix C: Full datasheet** from the link on the right of the webpage to a folder of your choice in your computer.

Problem 2/6: Load the dataset to Tableau [15 points]

1. Open Tableau. Load the Excel dataset. Select the sheet **Data** by double clicking on it.
2. Preview the dataset in Tableau Data Source. Observe the row header.

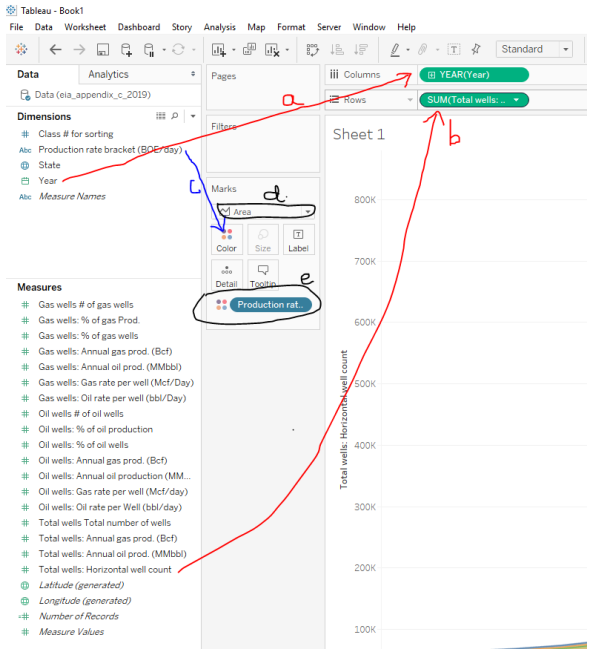


3. Check ‘Use Data Interpreter’ on the left side of the Data Source tab in Tableau. Observe the row header now.
4. Click on Manage Metadata- Red circle in the below image. Understand all the attributes. Change **Year** field from **Numeric** to **Date**.
5. What is the relation between Class # for Sorting and Production rate brackets? Change **Class # for Sorting** feature type to **String**.



Problem 3/6: Create the visualization in Figure 2 of the data source link **[20 points]**

1. Click on Sheet1. Let us try to create the visualization in Figure 2 in the dataset link.



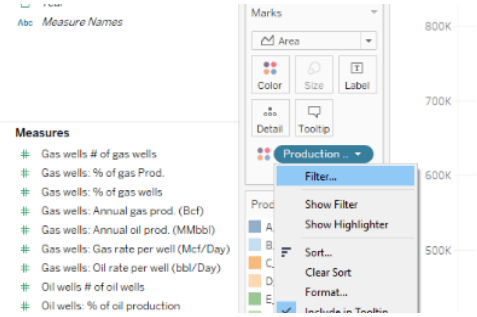
a. Select and drag **Year** from Dimensions to Columns.

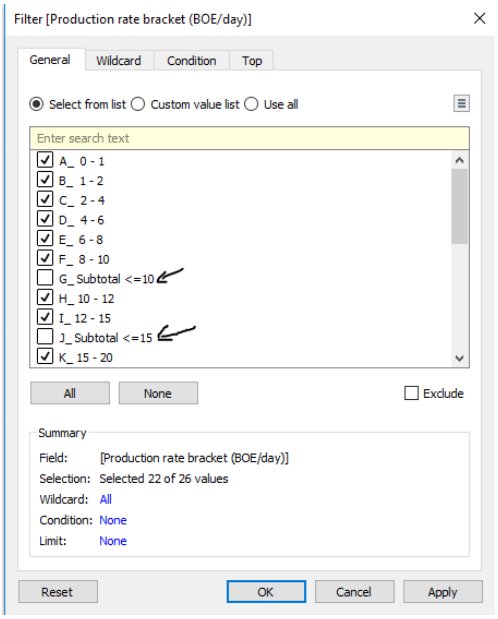
b. Select and drag **Total wells: Horizontal well count** to rows Select

c. Select **Production rate bracket (BOE/day)** and drag it on top of Colors

d. Click on pull down menu in Marks and select Area

e. Click **Production rate bracket (BOE/day)** field in Color and click filters. Unselect all Subtotal and Total fields as shown below. Click OK.





f. Drag the color legend to the left. Click on legend, click sort and select descending. Close the window by clicking on **x**

on ‘Total 

g. Double click Sheet1 and rename it as Hor Wells by Production Rate Brackets’

Complete the table below:

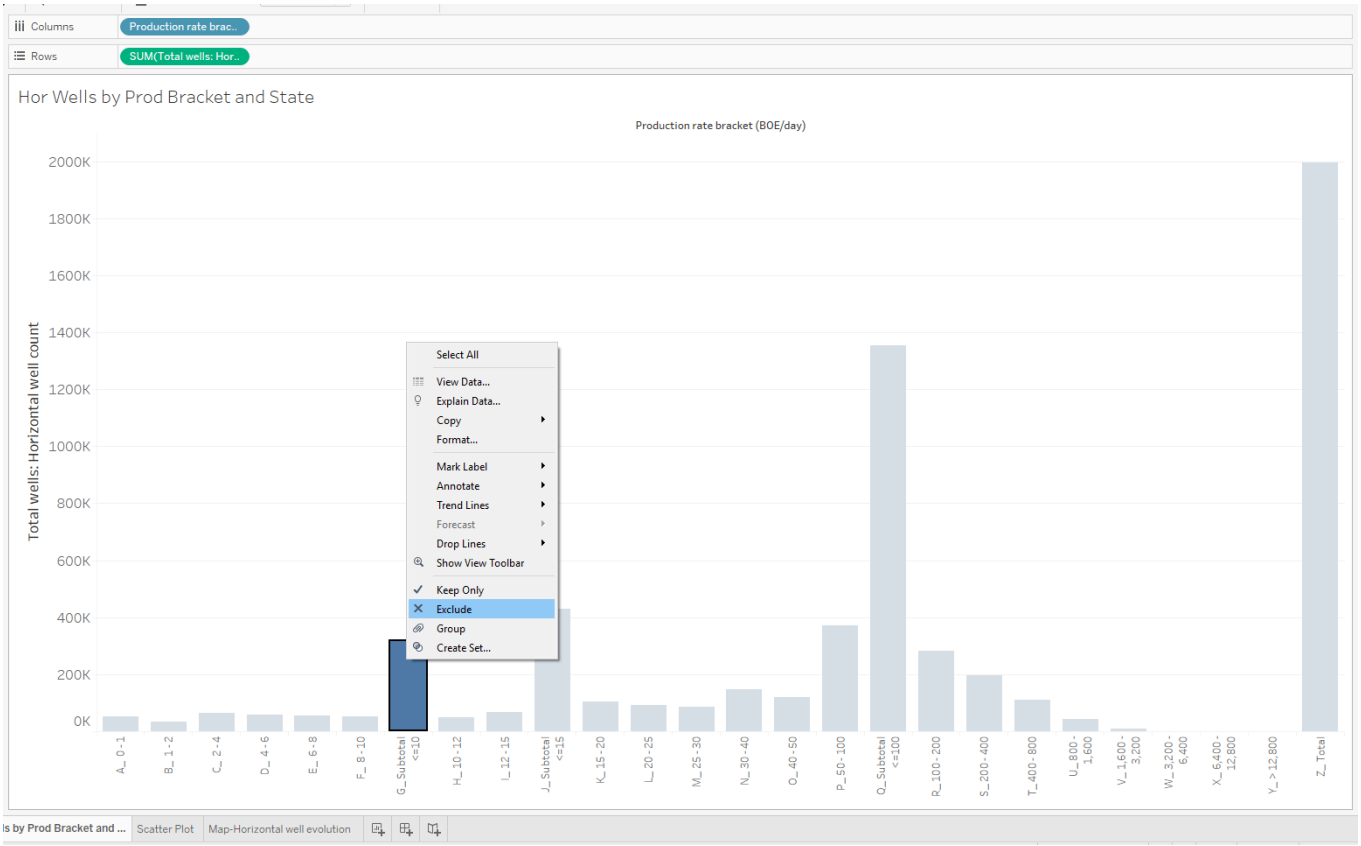
Is this viz same as Fig 2 on the EIA link? Choose one: Yes / No.	**No**
Why and what can you do to address the difference?	Different color schemes; not have to be the same color as long as data is accurate
List 3 observations from this visualization:	1. Total wells count increased annually from 2000 to 2020 2. Production rate P_50-100 bracket is largest compare to others annually 3. Y_>12,800 smallest of horizontal count, but product the highest of gas and oil
What type of visualization is this? What are the challenges in interpreting this visualization?	Stack/area chart Might not answer all questions such as making predictions

Problem 4/6: Visualize horizontal wells by production brackets and States using bar and line charts [20 points]

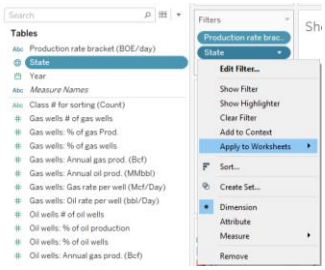
a. Add a new sheet and rename the sheet.
b. Add **Production rate bracket (BOE/day)** to columns and **Total wells: Horizontal well count** to rows.

What type of visualization do you see?	Bar Chart

c. Exclude the total categories as shown below:



d. **State** field to filters. Filter out **US.** Activate this filter for the related all sheets using this data source



What are the top 3 production brackets for horizontal wells?	*Hint: Use Sort* P_50-100, R_100-200, S_200-400

e. Add **State** field to color.

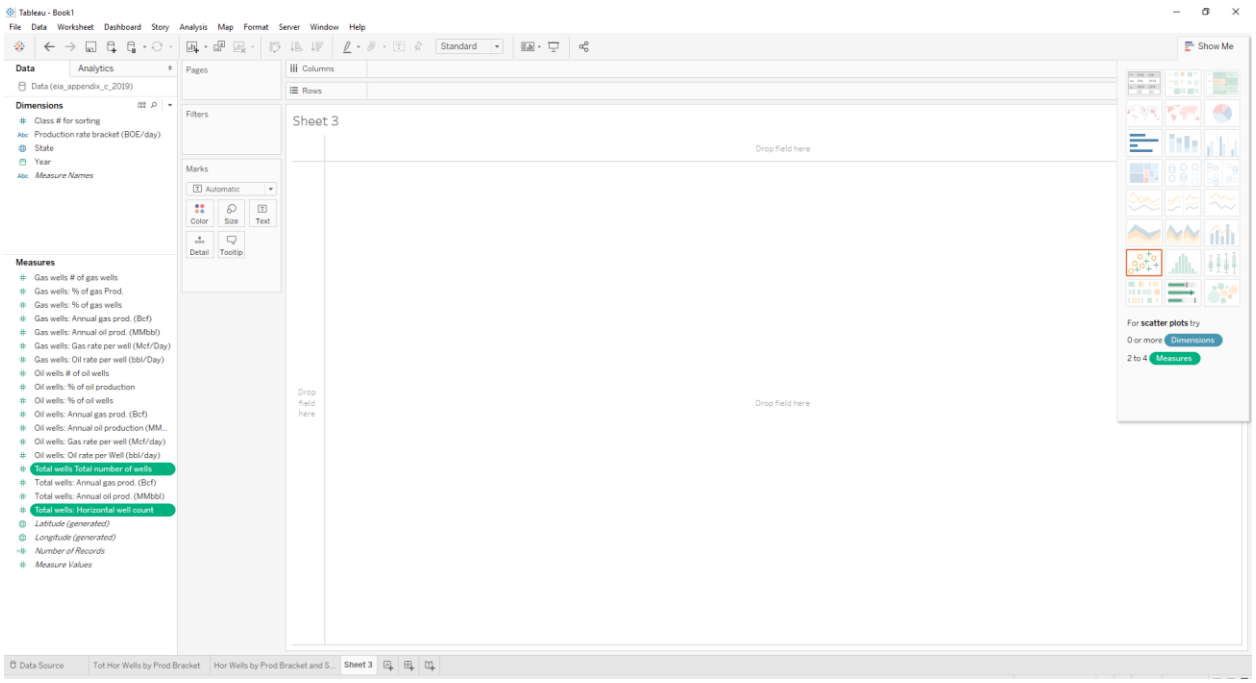
f. Change chart type to **Line** in the **Marks** card.

What is the leading state in Horizontal Wells and what is the highest production bracket for this state?	TX and prod bracket P_50-100
What is the highest production bracket range for ND, OK and PA?	ND: P_50-100 with 32,360 horizontals well count

	OK: P_50-100 horizontal well count with 23,314 PA:P_50-100 with 10,370 horizontal well count

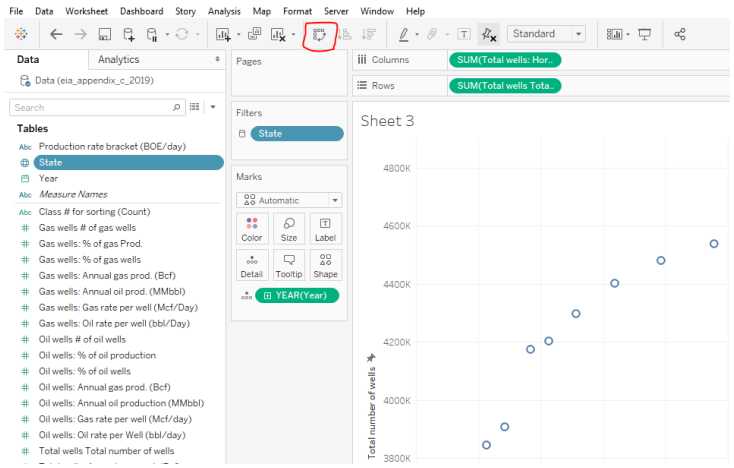
Problem 5/6: Visualize the relationship between Total Wells and Total Horizontal Wells over the years using a Scatter Plot [20 points]

a. Add a new sheet and rename the sheet.
b. Use Ctrl key to select **Total Wells: Total number of wells** and **Total Wells: Horizontal well count.**
c. Click Show me on the top right corner of the Tableau workspace

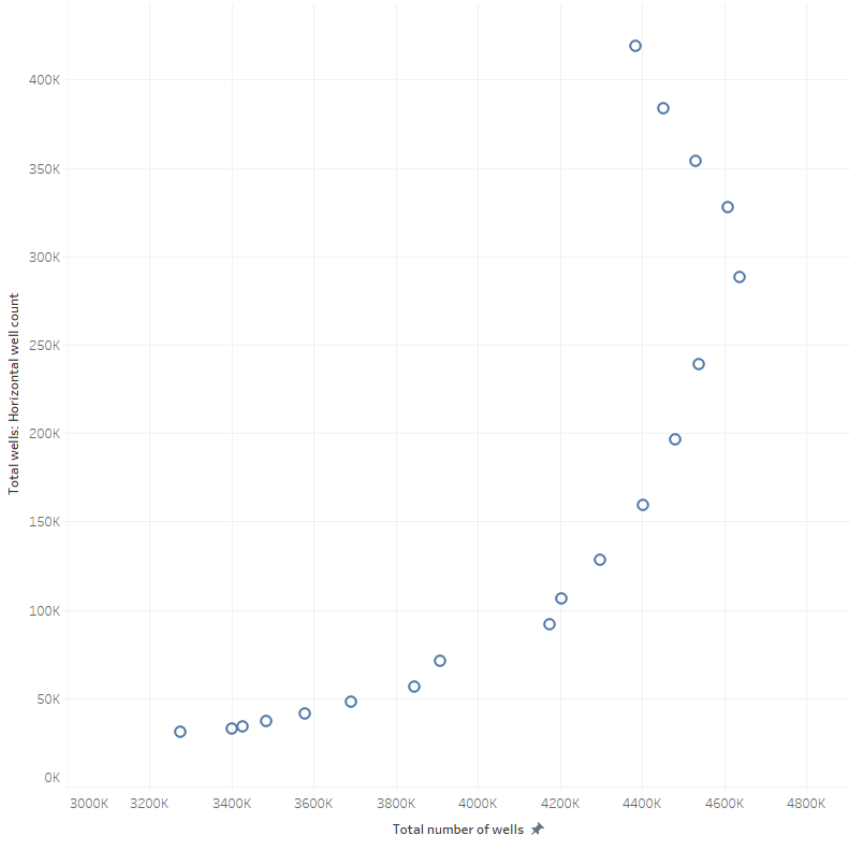


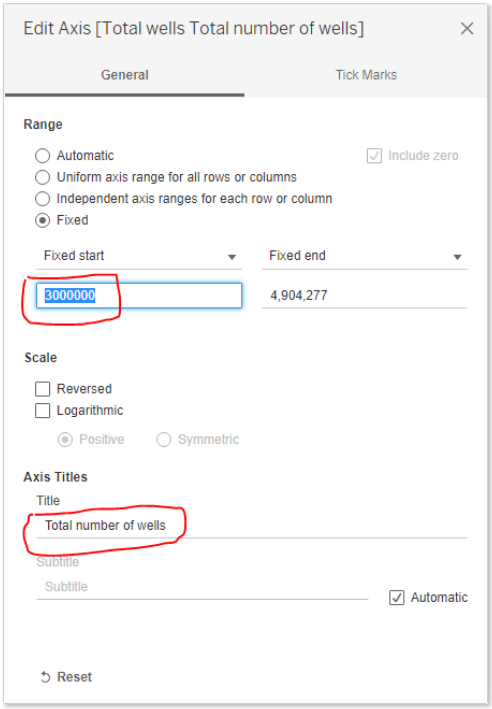
What is the recommended visualization?	Scatter plot

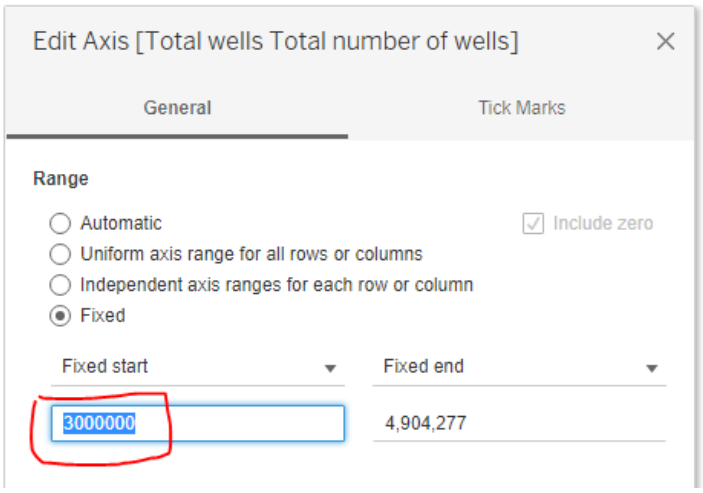
d. Click Scatter Plot on Show Me.
e. Select and drag **Year** field on Detail on the Marks Card.
f. Swap Rows and Columns- **Total Wells: Horizontal well count** to be on vertical axis (Y-axis).



g. Double click on the horizontal axis and change the axis range as shown below:





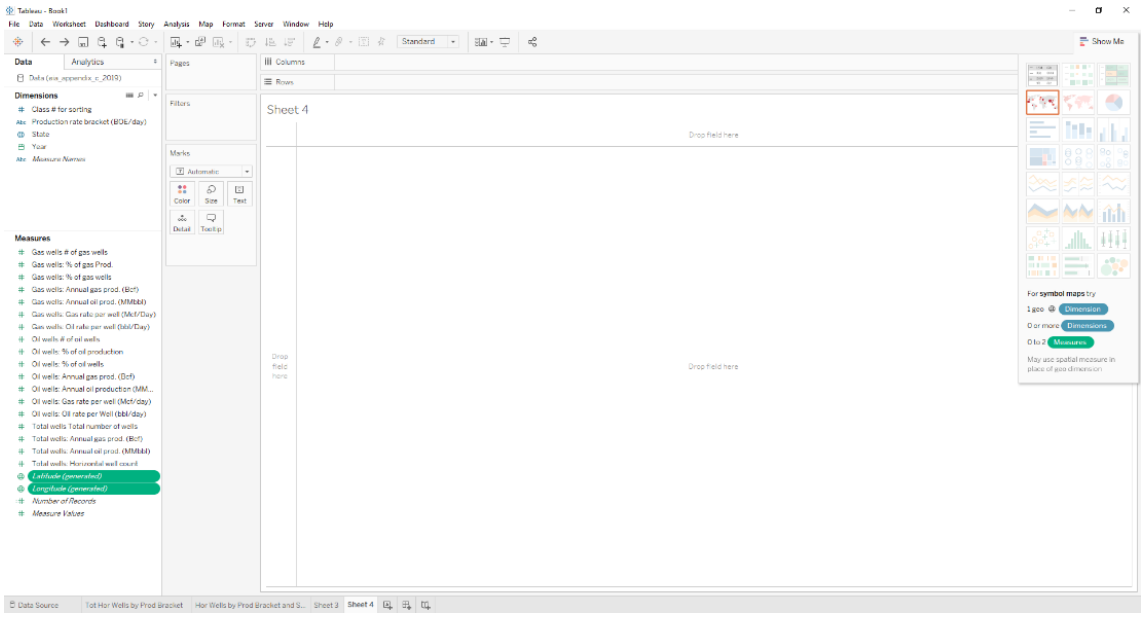


h. Drag **Year** to **Label** on the Marks card.

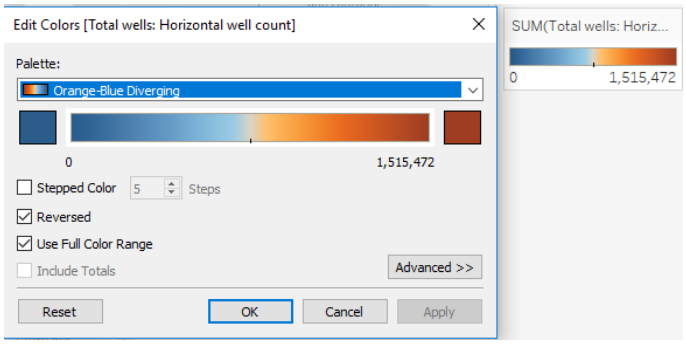
How do you describe the relationship? *(give full credit if any one of these or something similar is given as a response by the student)*	Total well count x total horizontal well count annually
What is the reason for the trend reversal?	Total well count decreased due to oil price crashed in 2014.

Problem 6/6: Visualize the States in a map and understand how Horizontal Well count evolved over time. [20 points]

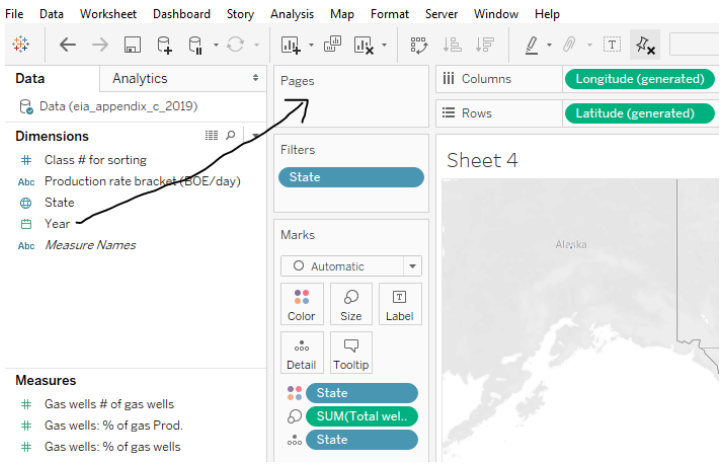
a. Add a new sheet and rename the sheet.
b. Use Ctrl key to select **Latitude** and **Longitude** and select **Symbol Maps** viz from Show Me.



c. Click and drag **States** field to Details on Marks Card.
d. Click and drag **Total Wells: Horizontal well count** to Size on Marks Card.
e. Click and drag **Total Wells: Horizontal well count** to Label.
f. Click and drag **Total Wells: Horizontal well count** to Color. Click on Color legend and select Palette as shown below. Select Reversed and Use Full Color Range.



g. Click and drag **Year** to Pages Card as shown below:



h. Play the Year animation to visualize evolution of Horizontal well count in each State over time (Years 2000-2018).

Complete the following table:

Horizontal wells in Arkansas in Year 2008?	3,008
Horizontal wells in Pennsylvania in Year 2012?	6,895
Horizontal wells in TX, OK and ND in Year 2014?	Tx: 135,717 OK: 35,773 ND: 27,351
Horizontal wells in TX, OK and ND in Year 2018?	TX: 197,107 OK: 48,562 ND: 41,460

i. TX Horizontal Well Count is much higher than other states. Filter out TX from State and replay the animation.