

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.

Tableau - Book1

File Data Server Window Help

Connections

eia_appendix_c_2019
Microsoft Excel

Sheets

☒ Use Data Interpreter
Data Interpreter might be able to clean your Microsoft Excel workbook.

brackets
Chart_of_pivot
Data
PivotTable
ReadMe
New Union

Data (eia_appendix_c_2019)

Connection: ☒ Live ☐ Extract Filters: 0 | Add

Data

Sort fields: Data source order Show aliases Show hidden fields 1,000 rows

State	Year	Production rate bracket (BOE/day)	Class # for sorting	Oil wells # of oil wells	Oil wells: % of oil wells	Oil wells: Annual oil production (MMbbl)	Oil wells: % of oil production	Oil wells: Oil rate per Well (bbl/day)	Oil wells: Annual gas prod. (Bcf)	Oil wells: Gas rate per well (Mcf/day)	Gas wells # of gas wells	Gas wells: % of gas wells	Gas wells: Annual gas prod. (Bcf)	Gas wells: % of gas Prod.	Gas wells: Gas rate per well (Mcf/Day)	Gas wells: Annual oil prod. (MMbbl)	Gas wells: Oil rate per well (bbl/Day)	Total wells Total number of wells	Total wells: Annual oil prod. (MMbbl)	Total wells: Annual gas prod. (Bcf)	Total wells: Horizontal well count
AK	2,018	A_ 0-1	1	10	0.520	0.001	0.000	0.41	0.002	0.71	17	3.640	0.002	0.000	0.54	0.0000	0.00	27	0.001	0.00	0.00
AK	2,018	B_ 1-2	2	3	0.160	0.001	0.000	1.32	0.001	1.46	7	1.500	0.014	0.000	8.20	0.0010	0.31	10	0.002	0.01	0.01
AK	2,018	C_ 2-4	3	16	0.840	0.013	0.010	2.62	0.012	2.28	7	1.500	0.034	0.010	16.49	0.0000	0.00	23	0.014	0.04	0.04
AK	2,018	D_ 4-6	4	12	0.630	0.010	0.010	4.42	0.006	2.65	5	1.070	0.036	0.010	25.09	0.0020	1.06	17	0.012	0.04	0.04
AK	2,018	E_ 6-8	5	9	0.470	0.016	0.010	6.32	0.012	4.51	2	0.430	0.009	0.000	20.84	0.0010	2.85	11	0.018	0.02	0.02
AK	2,018	F_ 8-10	6	11	0.580	0.022	0.010	7.20	0.031	10.12	6	1.280	0.039	0.010	35.05	0.0030	3.00	17	0.025	0.07	0.07
AK	2,018	G_ Subtotal <=10	6	61	3.200	0.064	0.040	3.92	0.063	3.87	44	9.420	0.134	0.030	13.28	0.0070	0.71	105	0.071	0.19	0.19
AK	2,018	H_ 10-12	7	7	0.370	0.022	0.010	9.29	0.031	13.00	2	0.430	0.019	0.000	48.54	0.0010	3.03	9	0.023	0.06	0.06
AK	2,018	I_ 12-15	8	19	1.000	0.066	0.040	11.23	0.072	12.36	6	1.280	0.136	0.030	62.00	0.0060	2.85	25	0.072	0.20	0.20
AK	2,018	J_ Subtotal <=15	8	87	4.560	0.152	0.100	6.17	0.166	6.76	52	11.130	0.289	0.070	22.80	0.0150	1.15	139	0.166	0.45	0.45
AK	2,018	K_ 15-20	9	30	1.570	0.140	0.090	14.76	0.150	15.87	10	2.140	0.194	0.050	85.20	0.0080	3.47	40	0.148	0.34	0.34
AK	2,018	L_ 20-25	10	33	1.730	0.215	0.140	20.43	0.148	14.08	13	2.780	0.530	0.130	126.18	0.0070	1.78	46	0.222	0.67	0.67
AK	2,018	M_ 25-30	11	20	1.050	0.126	0.080	24.50	0.092	18.01	12	2.570	0.673	0.170	154.79	0.0100	2.24	32	0.136	0.76	0.76
AK	2,018	N_ 30-40	12	55	2.880	0.578	0.370	30.89	0.438	23.40	9	1.930	0.666	0.170	204.73	0.0000	0.00	64	0.578	1.10	1.10

Go to Worksheet

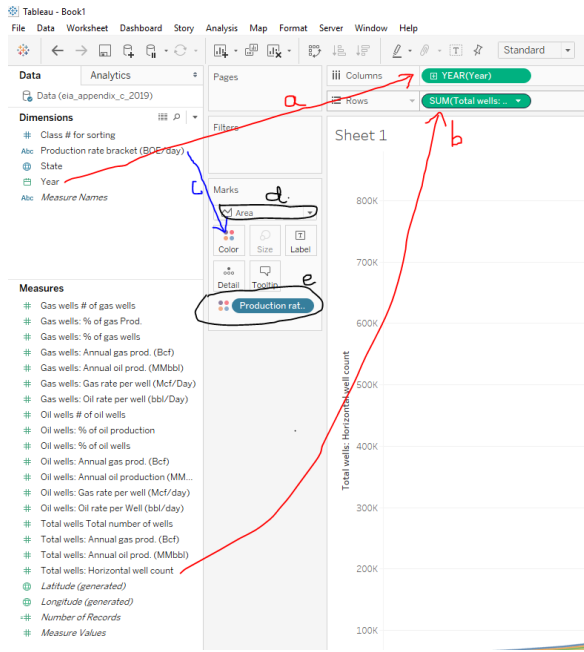
3. Check 'Use Data Interpreter' on the left side of the Data Source tab in Tableau. Observe the row header now.
4. Click on Manage Metadata- Red circle in the below image. Understand all the attributes. Change **Year** field from **Numeric** to **Date**.
5. What is the relation between Class # for Sorting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

Sort fields: Data source order

Field Name	Table	Remote Field Name
State	Data	State
Year	Data	Year
Production rate bracket (BOE/day)	Data	Production rate bracket (BOE/day)
Class # for sorting	Data	Class # for sorting
Oil wells # of oil wells	Data	Oil wells # of oil wells
Oil wells: % of oil wells	Data	Oil wells: % of oil wells
Oil wells: Annual oil production (MMbbl)	Data	Oil wells: Annual oil production (MMbbl)
Oil wells: % of oil production	Data	Oil wells: % of oil production
Oil wells: Oil rate per Well (bbl/day)	Data	Oil wells: Oil rate per Well (bbl/day)
Oil wells: Annual gas prod. (Bcf)	Data	Oil wells: Annual gas prod. (Bcf)
Oil wells: Gas rate per well (Mcf/day)	Data	Oil wells: Gas rate per well (Mcf/day)
Gas wells # of gas wells	Data	Gas wells # of gas wells
Gas wells: % of gas wells	Data	Gas wells: % of gas wells
Gas wells: Annual gas prod. (Bcf)	Data	Gas wells: Annual gas prod. (Bcf)
Gas wells: % of gas Prod.	Data	Gas wells: % of gas Prod.
Gas wells: Gas rate per well (Mcf/Day)	Data	Gas wells: Gas rate per well (Mcf/Day)
Gas wells: Annual oil prod. (MMbbl)	Data	Gas wells: Annual oil prod. (MMbbl)
Gas wells: Oil rate per well (bbl/Day)	Data	Gas wells: Oil rate per well (bbl/Day)
Total wells Total number of wells	Data	Total wells Total number of wells
Total wells: Annual oil prod. (MMbbl)	Data	Total wells: Annual oil prod. (MMbbl)
Total wells: Annual gas prod. (Bcf)	Data	Total wells: Annual gas prod. (Bcf)
Total wells: Horizontal well count	Data	Total wells: Horizontal well count

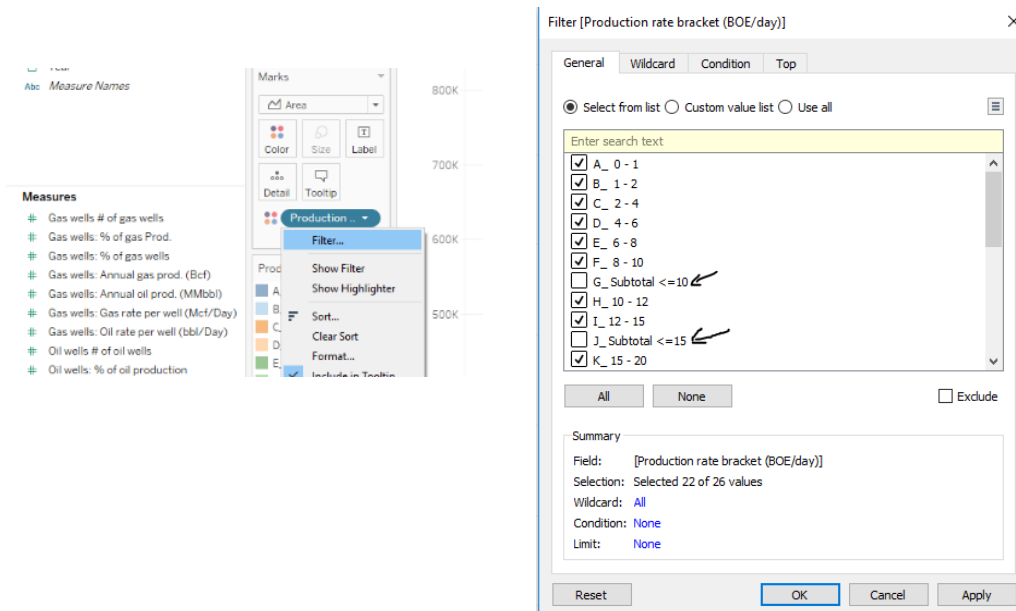
Problem 3/6: Create the visualization in Figure 2 of the data source link [20 points]

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



- Select and drag **Year** from Dimensions to Columns.
- Select and drag **Total wells: Horizontal well count** to rows
- Select **Production rate bracket (BOE/day)** and drag it on top of Colors
- Click on pull down menu in Marks and select Area
- Click **Production rate bracket (BOE/day)** field in Color and click filters. Unselect all Subtotal and Total fields as shown

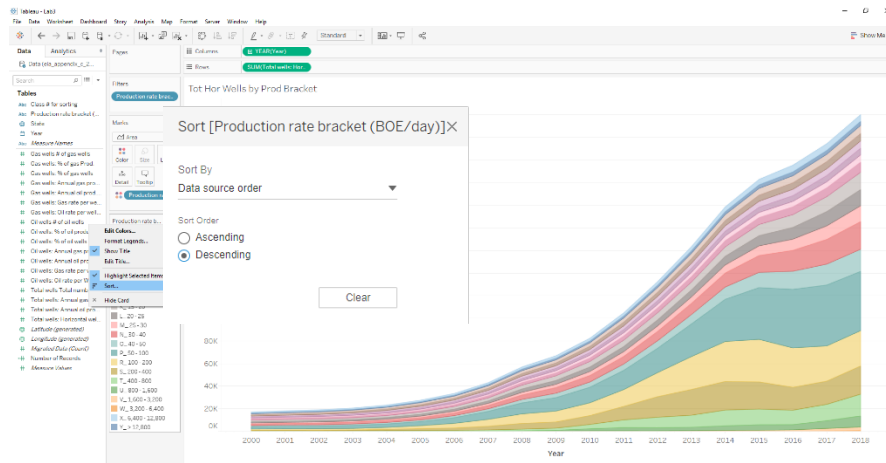
below. Click OK.



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

on
'Total

Complete the



g. Double click
Sheet1 and
rename it as
Hor Wells by
Production Rate
Brackets'
table below:

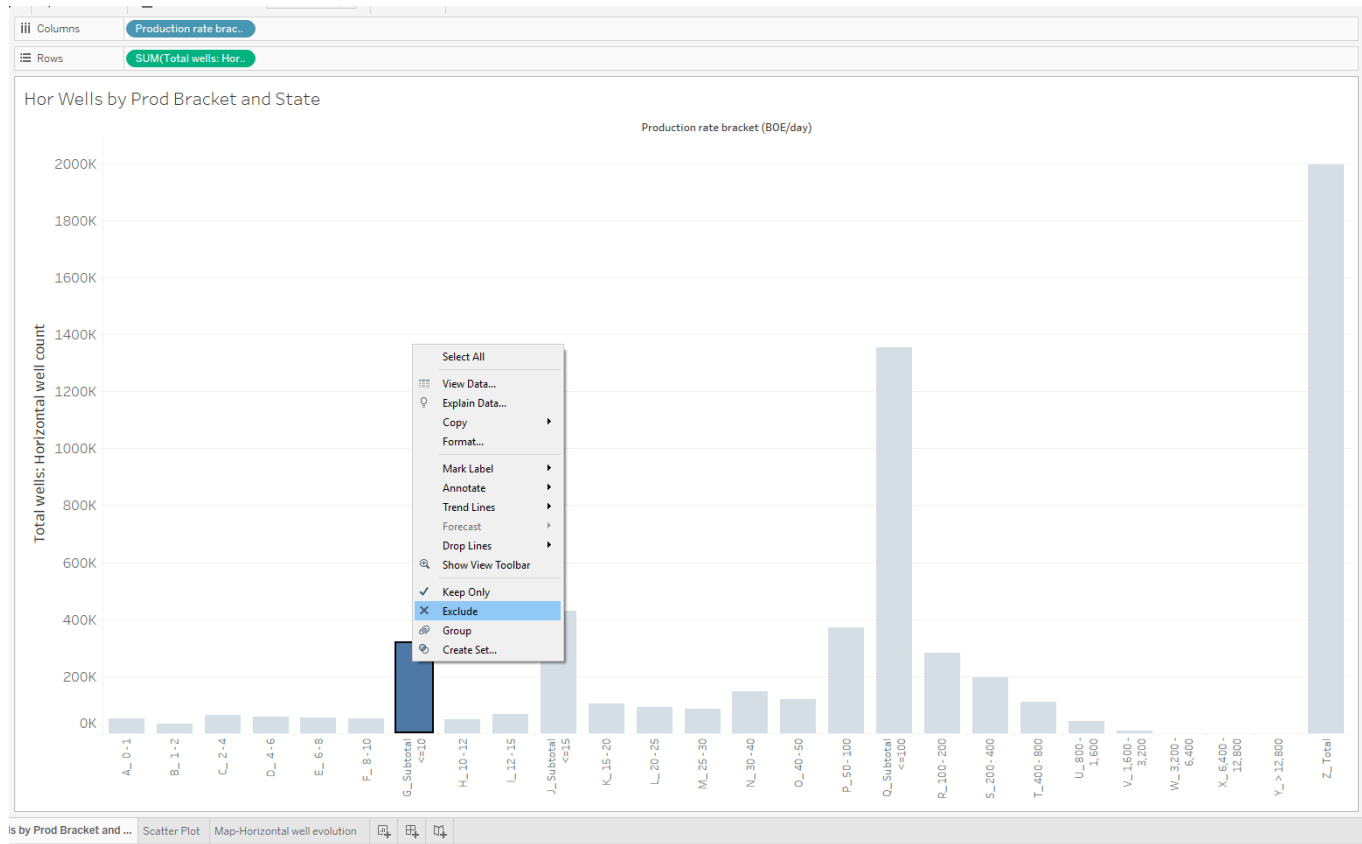
Is this viz same as Fig 2 on the EIA link? Choose one: Yes / No.	No
Why and what can you do to address the difference?	Different color schemes; not have to be the same color as long as data is accurate
List 3 observations from this visualization:	<ol style="list-style-type: none"> 1. Total wells count 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?	<p>Stack/area chart</p> <p>Might not answer all questions such as making predictions</p>

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

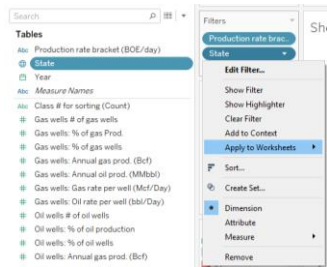
- a. Add a new sheet and rename the sheet.
- b. Add **Production rate bracket (BOE/day)** to columns and **Total wells: Horizontal well count** to rows.

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.

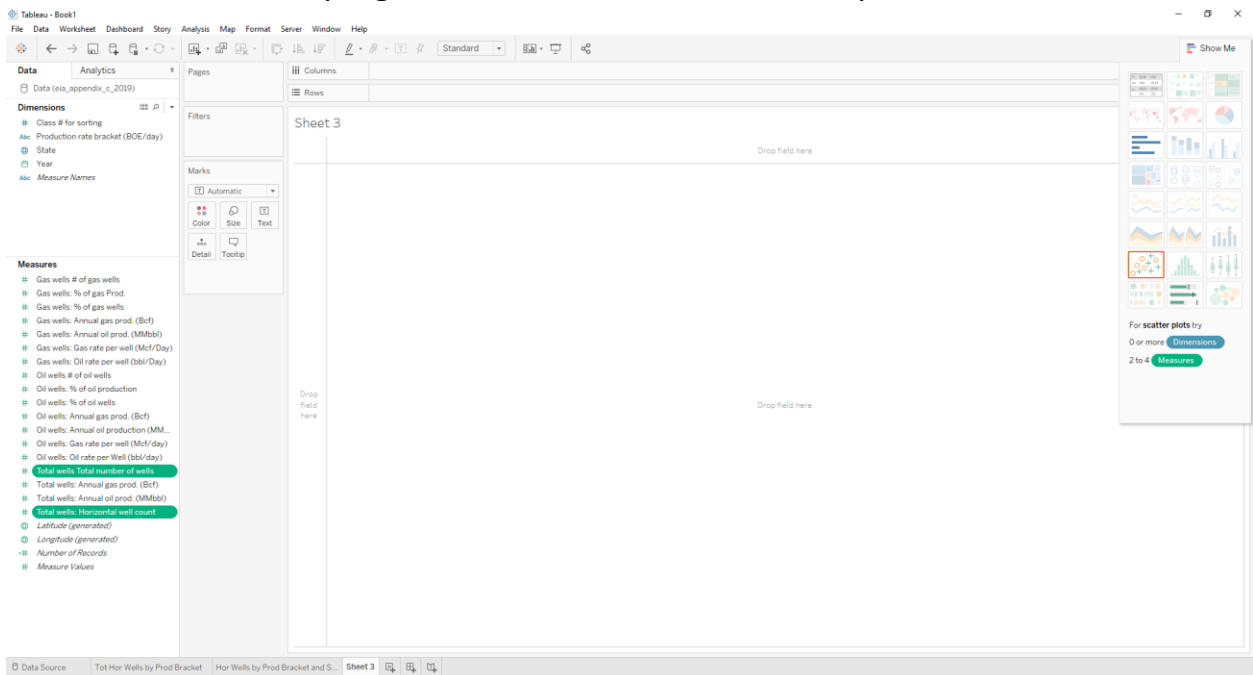
What is the leading state in Horizontal Wells and what is the highest production bracket for this state?	TX and prod bracket P_50-100
What is the highest production bracket range for ND, OK and PA?	ND: P_50-100 with 32,360 horizontals well count

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]

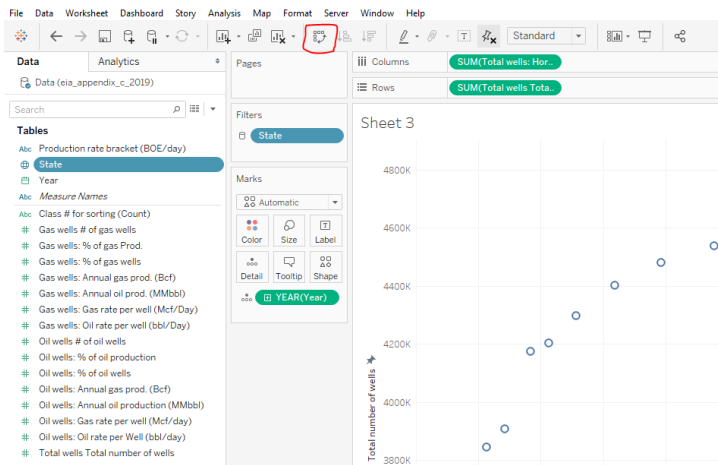
- Add a new sheet and rename the sheet.
- Use Ctrl key to select **Total Wells: Total number of wells** and **Total Wells: Horizontal well count**.
- Click Show me on the top right corner of the Tableau workspace



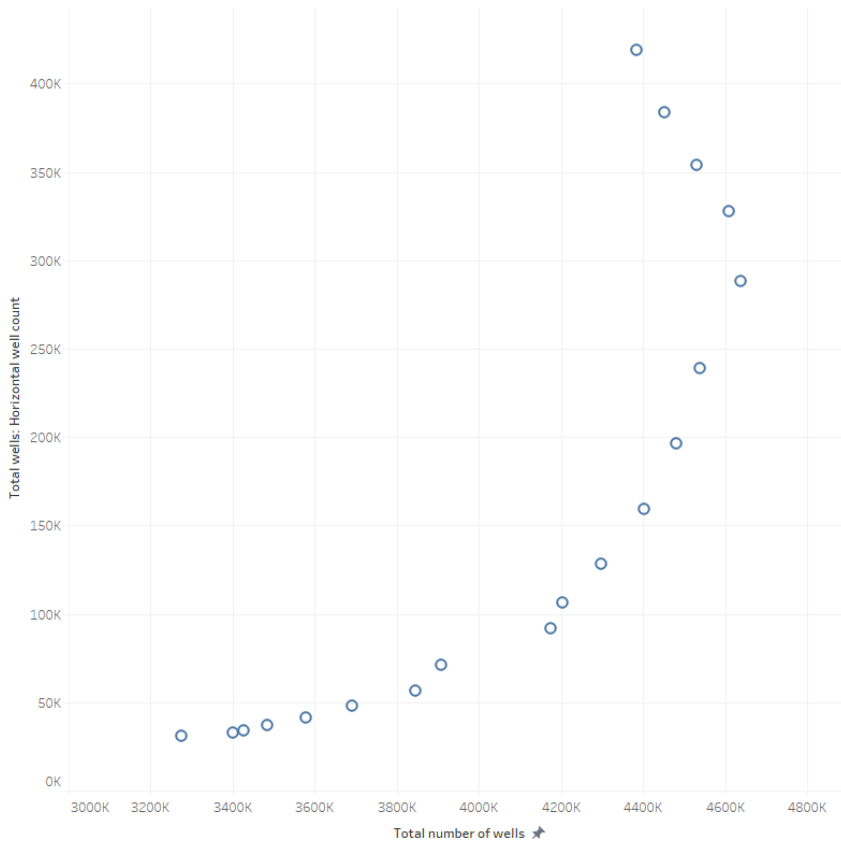
What is the recommended visualization?

Scatter plot

- Click Scatter Plot on Show Me.
- Select and drag **Year** field on Detail on the Marks Card.
- 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:



Edit Axis [Total wells Total number of wells]

General
Tick Marks

Range

☐ Automatic
☐ Uniform axis range for all rows or columns
☐ Independent axis ranges for each row or column
☒ Fixed

Fixed start

Fixed end

3000000

4,904,277

Scale

☐ Reversed
☐ Logarithmic

☒ Positive
☐ Symmetric

Axis Titles

Title

Total number of wells

Subtitle

Subtitle

☒ Automatic

Reset

Edit Axis [Total wells Total number of wells]

General
Tick Marks

Range

☐ Automatic
☐ Uniform axis range for all rows or columns
☐ Independent axis ranges for each row or column
☒ Fixed

Fixed start

Fixed end

3000000

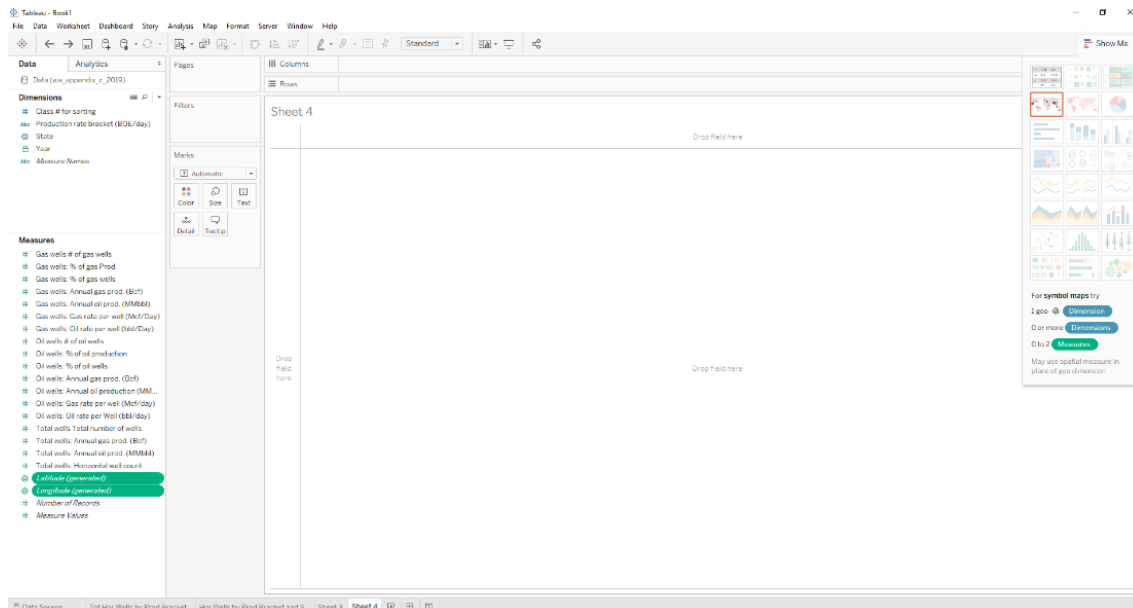
4,904,277

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

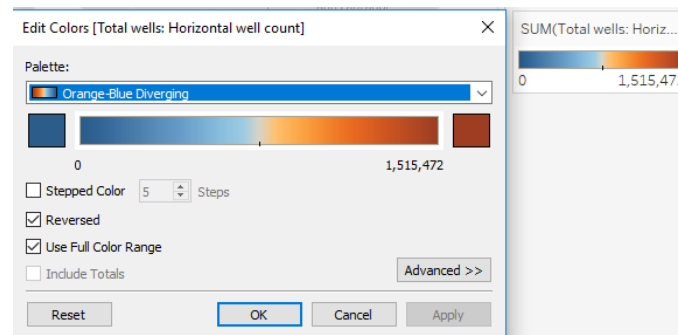
How do you describe the relationship? <i>(give full credit if any one of these or something similar is given as a response by the student)</i>	Total well count x total horizontal well count annually
What is the reason for the trend reversal?	Total well count decreased due to oil price crashed in 2014.

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

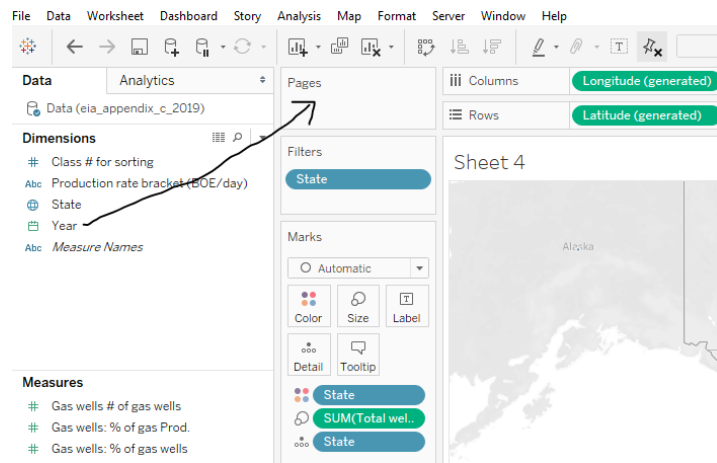
- Add a new sheet and rename the sheet.
- Use Ctrl key to select **Latitude** and **Longitude** and select **Symbol Maps** viz from Show Me.



- Click and drag **States** field to Details on Marks Card.
- Click and drag **Total Wells: Horizontal well count** to Size on Marks Card.
- Click and drag **Total Wells: Horizontal well count** to Label.
- 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.



- 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.