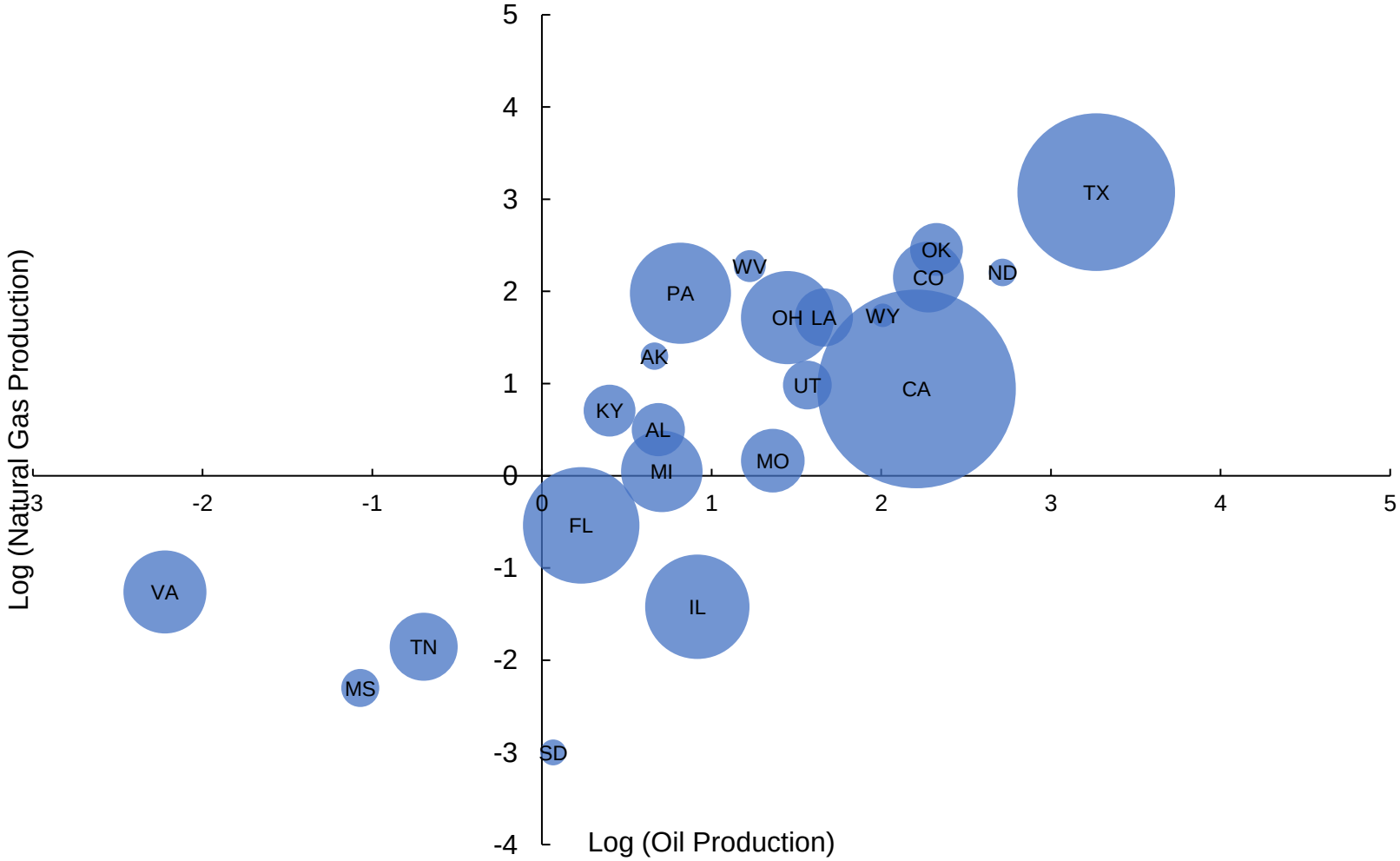
Visualizing Association

Figure 1. Log(Natural Gas Production), Log(Oil Production), and Real Gross Domestic Product by State, 2019



Source: *Author's elaboration with data from the Energy Information Administration*
Remark: *Bubble size denotes real GDP in 2012 USD.*

The relevant points of this data visualization are:

- The main purpose of this figure is to visualize association between log(natural gas production), log (oil production), and real GDP in 2019 using a bubble chart.
- As observed in Figure 1, there is a positive correlation between log(natural gas production) and log (oil production).
- The greater the log(natural gas production), the greater the log (oil production).
- Larger states in real GDP terms showed large log(natural gas production) and log (oil production), except for Florida, Virginia, and Tennessee.
- A logarithmic transformation was required to get rid of natural gas production and oil production.
- This type of analysis is an initial step to identify relevant factors that could affect the natural gas production prior to build an econometric model.
- This is an initial analysis to visualize the association between natural gas production with some potential drivers of its production such as oil production and real GDP; however, other factors might explain the natural gas production by state.

I presented in Table 1 the six Tufte's principles recommended for a data visualization along with a description of each principle. Additionally, I indicated which of these principles I used when creating Figure 1 and explained their applications.

Table1. Tufte's Principles applied in Figures 1

Tufte's Principles	Figure1	Detailed description	Application
Show comparisons	✓	Making comparisons is helpful in identifying magnitude visually	Figure 1 compares the log(natural gas production), the log(oil production), and the real GDP by state in 2019. The bubble size denotes the size of the economy of each state.
Show causality	✓	The data visualization could have as major goal to show causality. How a variable cause another visually could provide a first insight in identifying causation.	In general, Figure 1 shows a positive association between log(natural gas production), the log(oil production), and the real GDP by state. In order to identify statistical causation additional statistical techniques must be used.
Use multivariate data	✓	Use multiple variables to accomplish the goal of the data visualization	Figure 1 uses log(natural gas production), the log(oil production), and the real GDP by state in 2019.
Completely integrate text, images, and numbers	✓	Integrate relevant notes, remarks, and images to better inform your audience about the information you want to communicate	Figure 1 contains labels of each of the states involved in the analysis; the size of the bubble refers to the size of the economy of each state.
Establish credibility	✓	One form to establish credibility is to include the data source and also start from the origin or allow the reader to identify the scale you are using	I included the source at the end of Figure 1.
Focus content	✓	Use space efficiently and avoid *chartjunk* by including relevant data to communicate your idea and stress on your main goal of your data visualization.	The transparency of each bubble aids to avoid visual overcrowding when two or more bubbles overlap. The scale in the axis aids to visualize the positive association between the variables.