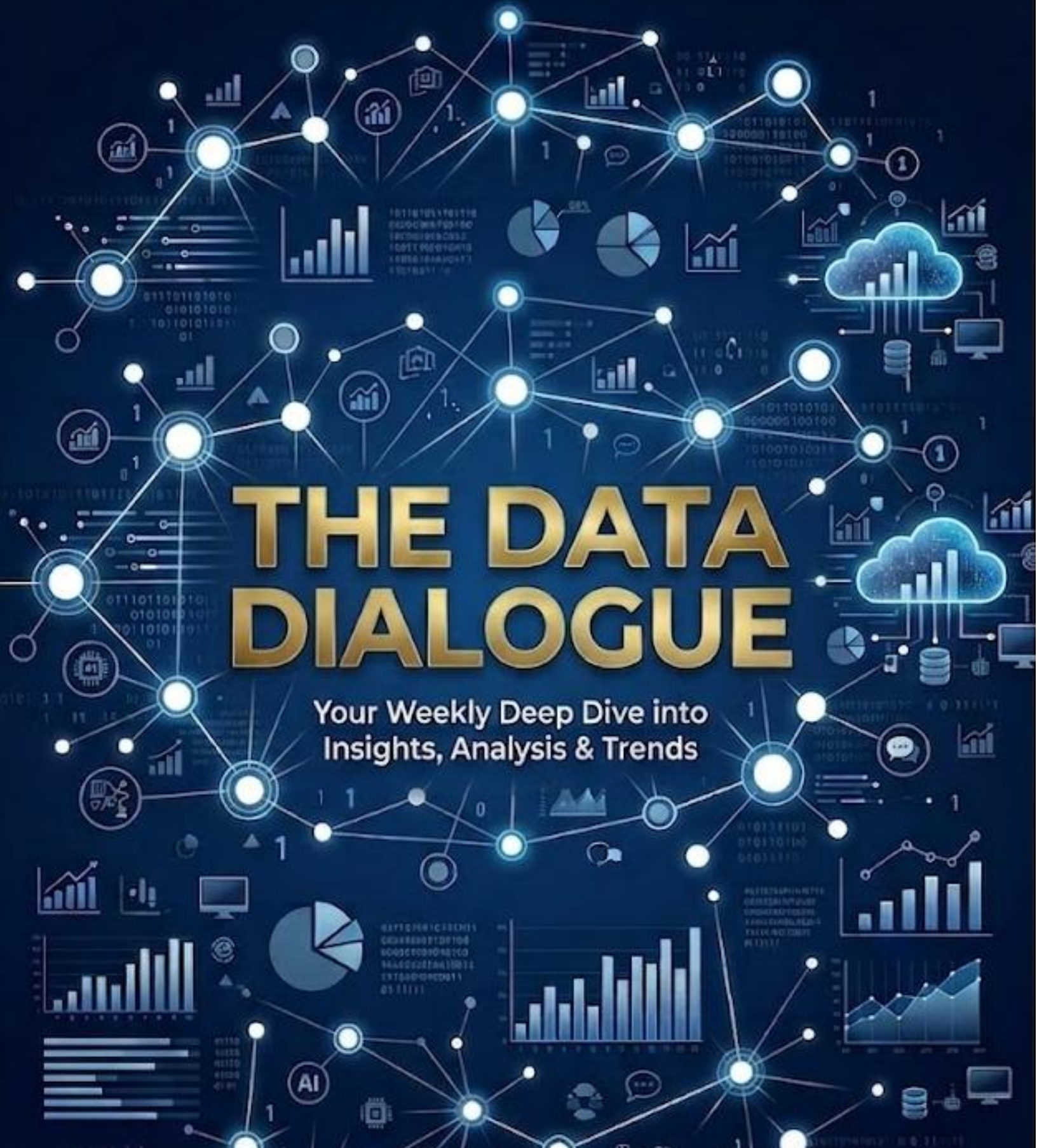


VOLUME 1| ISSUE 4| August 2026

THE DATA DIALOGUE

Your Weekly Deep Dive into
Insights, Analysis & Trends



The Data Dialogue.

Simplified. Not Simplistic

August 5, 2026 | Vol.1 | Issue 4 | Stepwise Regression

In-Focus

Intellectual Flexibility: The Unspoken Skill

I am sure most of you, learners or instructors are familiar with the following situation: a teacher adopts a specific teaching method believing it is the best. After the first quiz, he/she finds out that the majority of students' grades are very low and only handful of students achieved good grades. The results came as a shock for the teacher who expected the grades to be higher or at least variable enough. Now, as a teacher, he/she would go down in one of two ways: blaming students for poor performance or deciding to change his/her teaching technique (the example can go the other way with a learner's study method).

Being wary of it or not, if the teacher decides to change his/her teaching technique (a data-driven decision), he/she has already made a decision to listen to data and revisit their theory about teaching. This ability to interpret results and revise one's initial position (hypothesis) is what I call "intellectual flexibility". I strongly believe it is the most important skill any researcher might have, and the most unspoken of. It is a meta-skill allowing who wields it to learn and perform other skills more intelligently.

If you ask how to acquire it, the answer is by practicing your craft more & sharing your work so that others can spot your mistakes, and you can learn from them...and repeat. Reading a book, receiving formal education, or sitting in a lab will get you nowhere as long as you are at the receiving end of the chain. Your intellectual flexibility is going to take shape and be honed once you roll up your sleeves and start doing the heavy lifting yourself. You stumble, share it with peers who are able to help you and you turn it into experience engraved in your mind.

So, next time when interpreting your data, use all your intellectual tools and imagination to see its variation (subsets, patterns and sub-categories patterns, outliers, implications, ...etc) to catch up on what you miss and what the data tries to tell you. You will learn to listen by doing; not by pausing.

The Data Dialogue.

Simplified. Not Simplistic

August 5, 2026 | Vol.1 | Issue 4 | Stepwise Regression

Tools

Dplyr () in R

- **One of the most important libraries to learn and work with in R.**
- Its main advantage is readability of its data management commands. In dplyr, you can simply ask R to summarise(), filter (), arrange(), mutate (); a direct, intuitive command in a verb form. Before dplyr one had to use indexing format that works on the data frame. More interestingly, you can chain two verbs/commands together using the idea of “pipe”, where the main connector is the operator (%>%) or (|>). It made the code documentation easier to write, debug and share with others. It also works perfectly with live streaming databases.

Pitfall Patrol



Using dplyr() when working with large datasets



Use dplyr() when working with live-streaming databases and small dataset.

Use data.table() when working with large dataset.

Why?

Unlike dplyr(), data.table () does not create a copy of the dataset in the memory of R. It means that computer performance is optimized.

The Data Dialogue.

Simplified. Not Simplistic

August 5, 2026 | Vol.1 | Issue 4 | Stepwise Regression

Pandora's Box *Stepwise Regression: A Cooking Lesson?!*

Building a model requires selecting a number of variables to include in this model. Adding more or fewer variables is like adding too much or too little salt to a meal while preparing it. Neither will make it taste good. So, analysts came up with an automated way that allow the computer to build the model by adding a variable every step and measure the significance it adds or retracts from the model (known as forward selection technique) or adding all the variables in the beginning and then removes one variable every step and measures how this removal affects (increase or decrease) the strength of the model, known as the backward elimination technique. Seems easy, quick, automated and intuitively removes the redundant variables (in technical terms reducing multicollinearity), right? Indeed- but nothing is cost-free!

Stepwise regression is criticized for being of lower predictive power, especially in machine learning domain. Think of this: while preparing your meal, you kept adding a few grains of salt until you reached the perfect taste for "you", but you could not know whether this level of salt tastes

good for your guests or clients. So, your training dataset cannot be used to test future datasets. As a consequence, the measurement is very unstable because if you tamper a little bit by adding another pinch of salt, everything changes. Stepwise regression is tainted with instability that results from its sensitivity to slight changes in data. Finally, if two variables have the same value, it removes one without any theoretical/ domain specific reason for it.

To overcome these concerns, statisticians developed a number of alternatives and the most important ones (and I believe it to be more used compared to others) are penalized regression techniques such as Lasso regression. Lasso regression is like a master chef who punishes you for building a complex model by adding too many variables. That's: "did you just add too much salt?", says Lasso regression, "You lost a point of your score. Keep doing this and you will lose more points and get kicked out of the kitchen eventually". This penalty forces you to keep the model simple and efficient.

The Data Dialogue.

Simplified. Not Simplistic

August 5, 2026 | Vol.1 | Issue 4 | Stepwise Regression

Before you Go

Did it happen that you made a mistake while conducting data analysis? If so, send us a punchy 250 words describing this mistake and how you fixed it.

If it's humanized, we'll publish it next week with your name, title and link to your profile.

Disclaimer: "We reserve the right to edit all submissions for conversational tone, length, and formatting to ensure it matches the Data Dialogue voice"

You will find the submission button on my website.