**Three R Libraries Every Data Scientist Should Know (Even if You Use Python)**

**Powerful R libraries built by the World’s Biggest Tech Companies**



Photo by [Denis Pavlovic](https://unsplash.com/@itsdenispavlovic?utm_source=unsplash&utm_medium=referral&utm_content=creditCopyText) on [Unsplash](https://unsplash.com/s/photos/coder?utm_source=unsplash&utm_medium=referral&utm_content=creditCopyText)

**Be sure to** [**subscribe here**](https://terenceshin.medium.com/membership) **and to my** [**personal newsletter**](https://terenceshin.substack.com/embed) **to never miss another article on data science guides, tricks and tips, life lessons, and more!**

**Introduction**

For the longest time, I was quite against using R for no other reason other than the fact that it wasn’t Python.

But after playing around with R for the past few months, I realized that R outclasses Python in several use cases, particularly for statistical analyses. As well, R has some powerful packages that were built by the world’s biggest tech companies, and they **aren’t** in Python!

And so, in this article, I wanted to go over three R packages that I **highly** **recommend** that you take the time to learn and equip in your arsenal of tools because they are seriously powerful tools.

Without further ado, here are three R packages that every data scientist should know, EVEN IF YOU ONLY USE PYTHON:

1. Causal Impact w/ Google
2. Robyn w/ Facebook
3. Anomaly Detection w/ Twitter

**Be sure to** [**subscribe here**](https://terenceshin.medium.com/membership) **and to my** [**personal newsletter**](https://terenceshin.substack.com/embed) **to never miss another article on data science guides, tricks and tips, life lessons, and more!**

**1. Causal Impact (Google)**

**[CausalImpact](https://google.github.io/CausalImpact/CausalImpact.html" \t "_blank)**

**[What does the package do? This R package implements an approach to estimating the causal effect of a designed…](https://google.github.io/CausalImpact/CausalImpact.html" \t "_blank)**

[google.github.io](https://google.github.io/CausalImpact/CausalImpact.html" \t "_blank)

Let’s say your company launched a new TV ad for the Super Bowl and they wanted to see how it impacted conversions. Causal impact analysis attempts to predict what would have happened if the campaign never occurred — this is called the counterfactual.

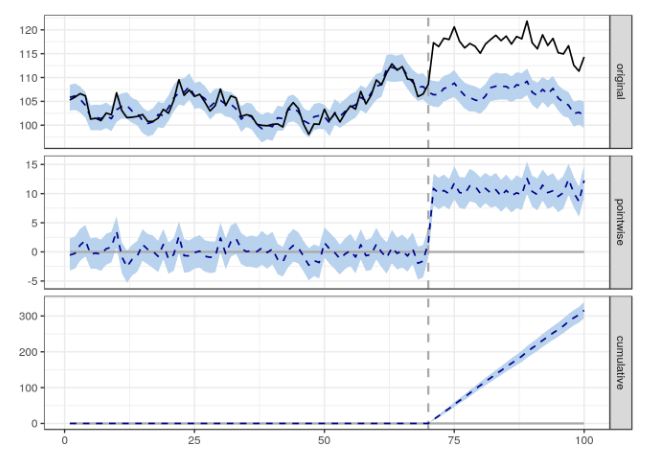


Image created by author

To give an actual example of what Causal impact does, it attempts to predict the counterfactual, i.e. the blue dotted line in the top graph, and then it compares the actual values to the counterfactual to estimate the delta.

Causal impact is super useful for marketing initiatives, expanding to new regions, testing new product features, and more!

**2. Robyn (Facebook)**

**[Robyn](https://facebookexperimental.github.io/Robyn/" \t "_blank)**

**[Robyn is an automated Marketing Mix Modeling (MMM) code. It aims to reduce human bias by means of ridge regression and…](https://facebookexperimental.github.io/Robyn/" \t "_blank)**

[facebookexperimental.github.io](https://facebookexperimental.github.io/Robyn/" \t "_blank)

Marketing Mix Modelling is a modern technique used to estimate the impact of several marketing channels or campaigns on a target variable, like conversions or sales.

Marketing Mix Models (MMMs) are extremely popular, more than attribution models, because they allow you to measure the impact of immeasurable channels like TV, billboards, and radio.

Typically, Marketing Mix Models take months to build from scratch. But Facebook created a new R package, called Robyn, that can create a robust MMM in minutes.

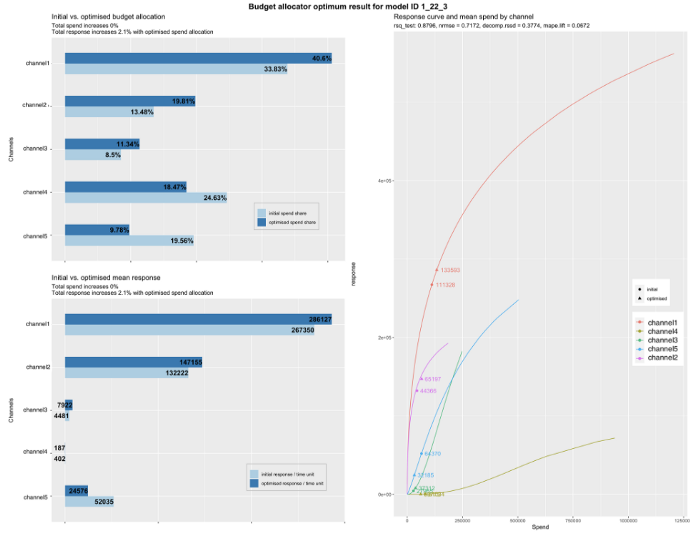


Image created by author

Not only can you assess the effectiveness of each marketing channel with Robyn, but you can also optimize your marketing budget with it too!

**Be sure to** [**subscribe here**](https://terenceshin.medium.com/membership) **and to my** [**personal newsletter**](https://terenceshin.substack.com/embed) **to never miss another article on data science guides, tricks and tips, life lessons, and more!**

**3. Anomaly Detection (Twitter)**

**[GitHub - twitter/AnomalyDetection: Anomaly Detection with R](https://github.com/twitter/AnomalyDetection" \t "_blank)**

**[AnomalyDetection is an open-source R package to detect anomalies which is robust, from a statistical standpoint, in the…](https://github.com/twitter/AnomalyDetection" \t "_blank)**

[github.com](https://github.com/twitter/AnomalyDetection" \t "_blank)

Anomaly detection, also known as outlier analysis, is a method that identifies data points that differ significantly from the rest of the data.

A subset of general anomaly detection is **anomaly detection in time-series data,** which is a unique problem because you have to consider the trend and seasonality of the data as well.

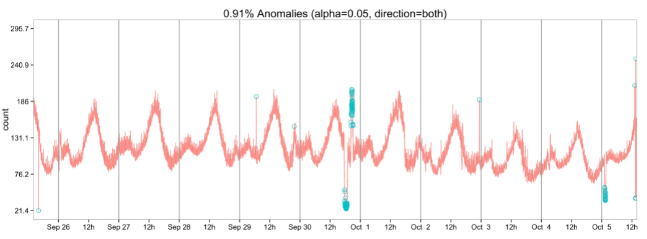


Image created by author

Twitter solved this problem by creating an anomaly detection package that does it all for you. It’s an intricate algorithm that can identify global and local anomalies. Aside from time series, it can also be used to detect anomalies in a vector of values.