Visualisation alone is not enough to solve most data analysis challenges. The data may be too big or too messy to show in a single plot. In this talk, I'll outline my current thinking about how the synthesis of visualisation, modeling, and data manipulation allows you to effectively explore and understand large and complex datasets.

There are three key ideas:

1. Using tidyr to make nested data frame, where one column is a list of data frames.
2. Using purrr to use function programming tools instead of writing for loops
3. Visualising models by converting them to tidy data with broom, by David Robinson.

This work is embedded in R so I'll not only talk about the ideas, but show concrete code for working with large sets of models. You'll see how you can combine the dplyr and purrr packages to fit many models, then use tidyr and broom to convert to tidy data which can be visualised with ggplot2.

Hadley is Chief Scientist at RStudio and a member of the R Foundation. He builds tools (both computational and cognitive) that make data science easier, faster, and more fun. His work includes packages for data science (ggplot2, dplyr, tidyr), data ingest (readr, readxl, haven), and principled software development (roxygen2, testthat, devtools). He is also a writer, educator, and frequent speaker promoting the use of R for data science.

Category

[Science & Technology](https://www.youtube.com/channel/UCiDF_uaU1V00dAc8ddKvNxA)

License

Standard YouTube License