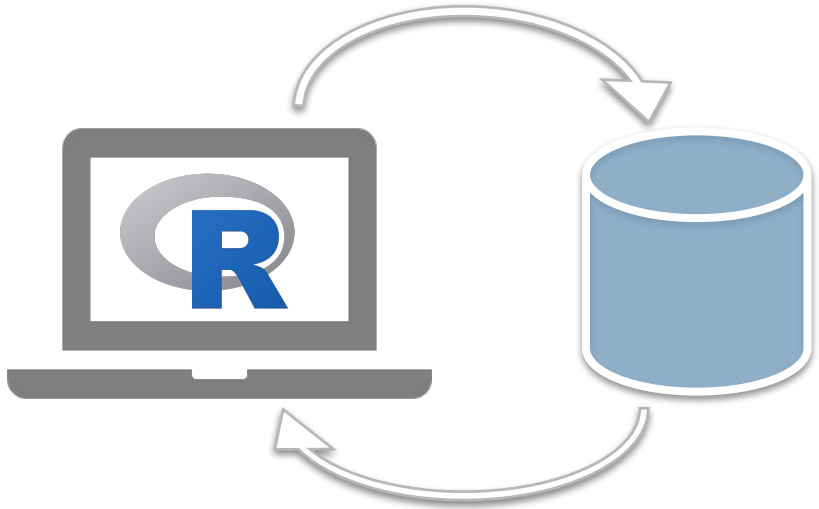
Modeling in Databases with R

Edgar Ruiz

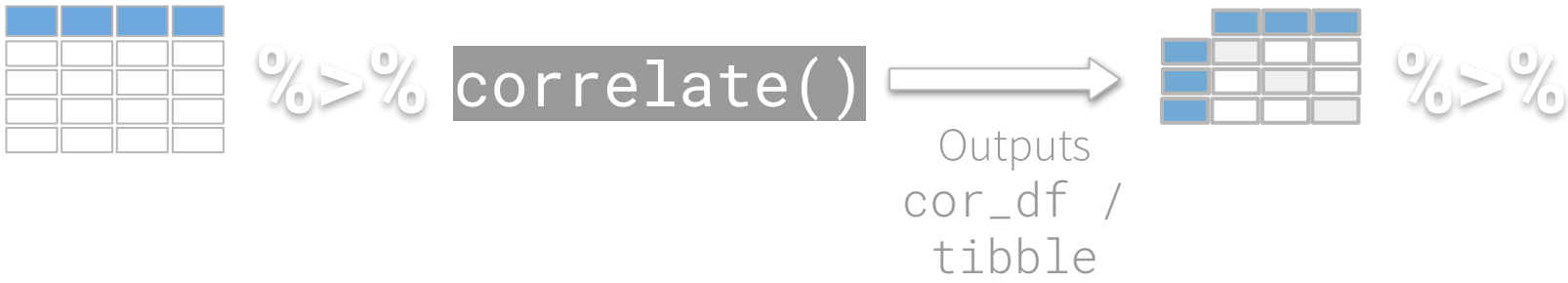
theotheredgar

July 2019

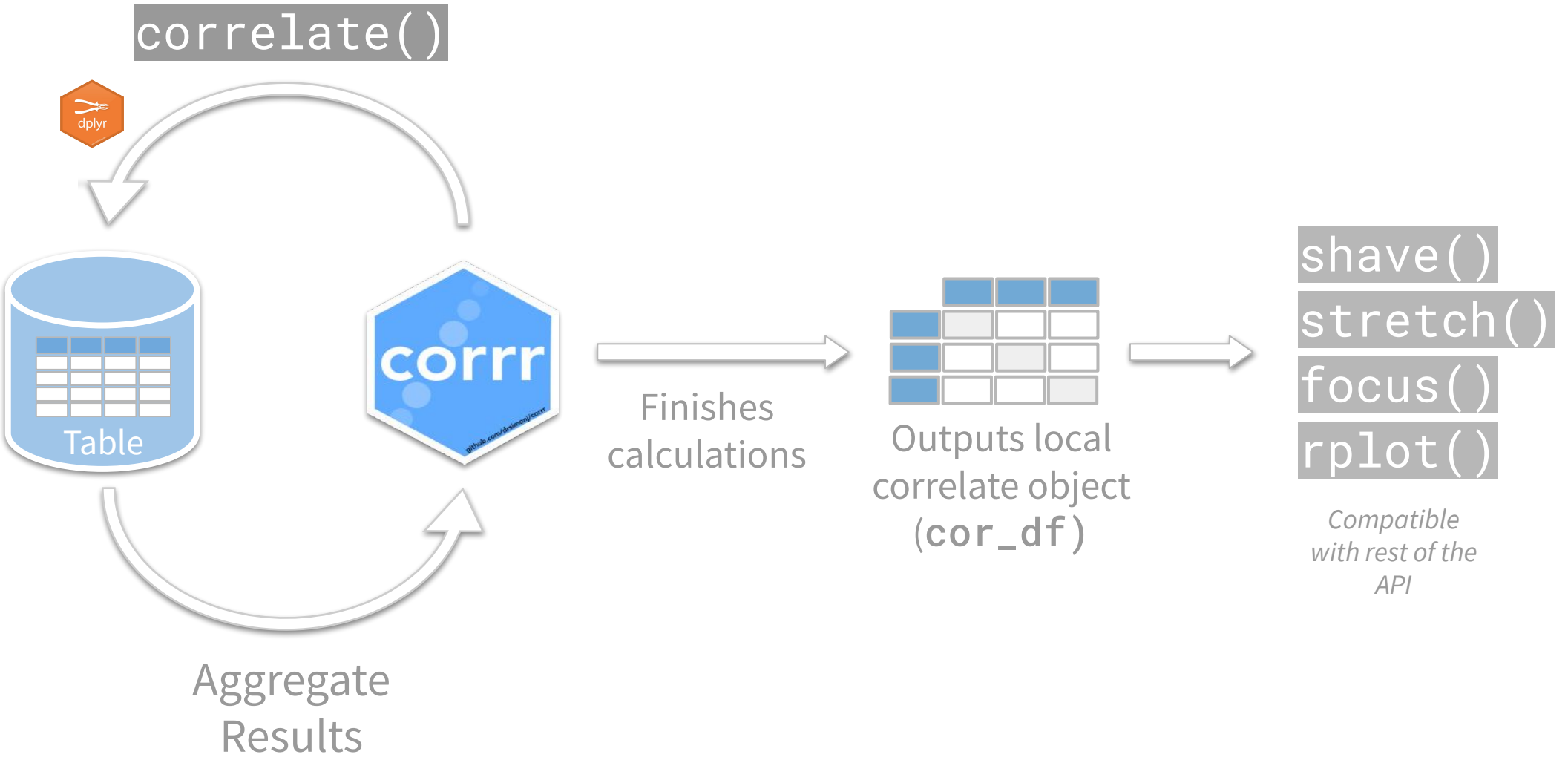


corrr

corrr package

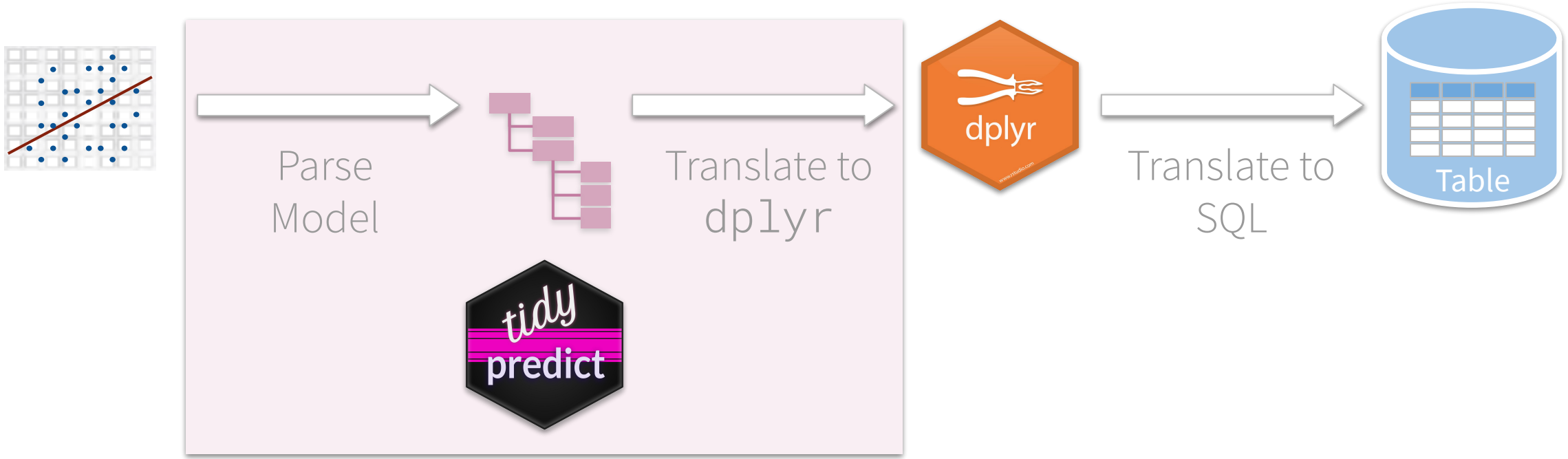


corrr database integration

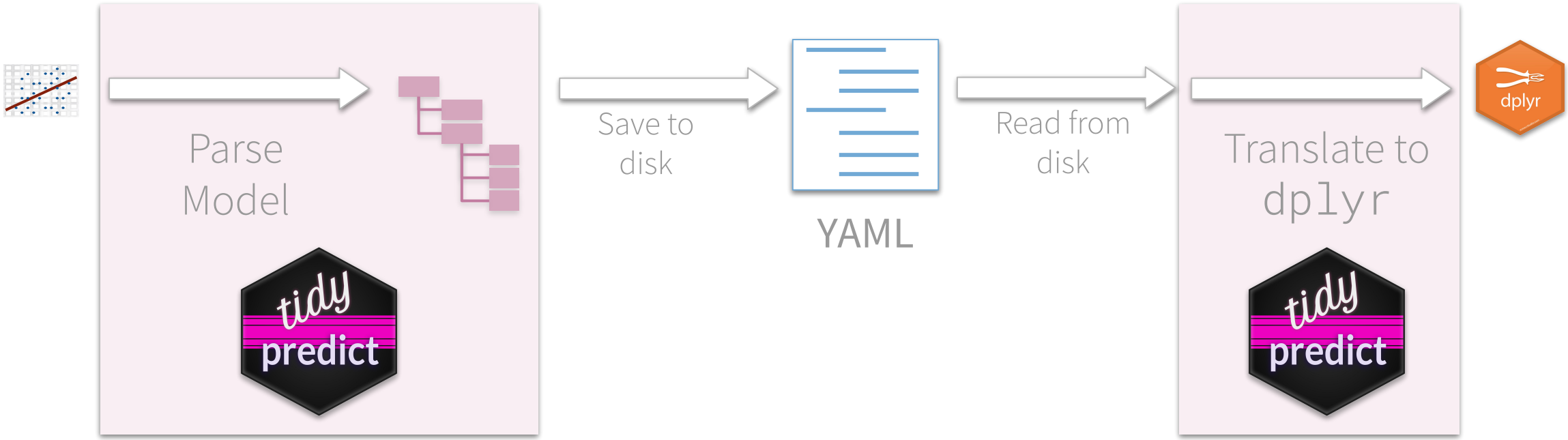


tidypredict

`tidypredict` workflow

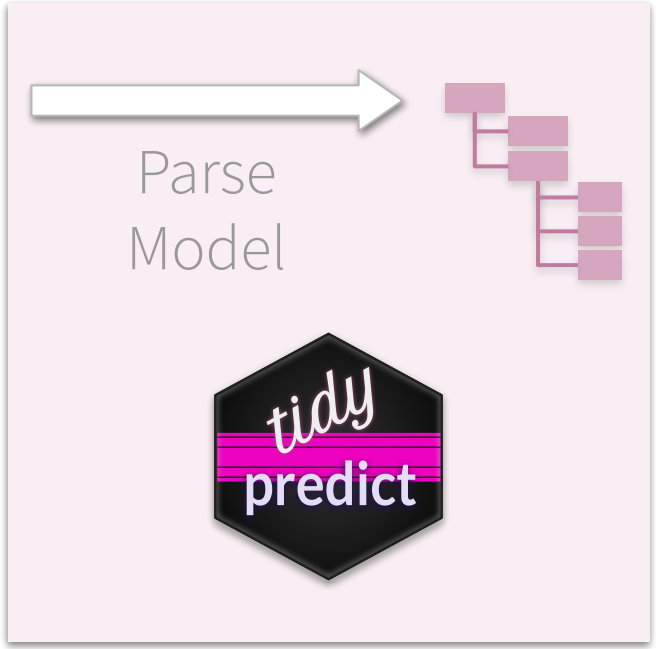


Save the parsed model



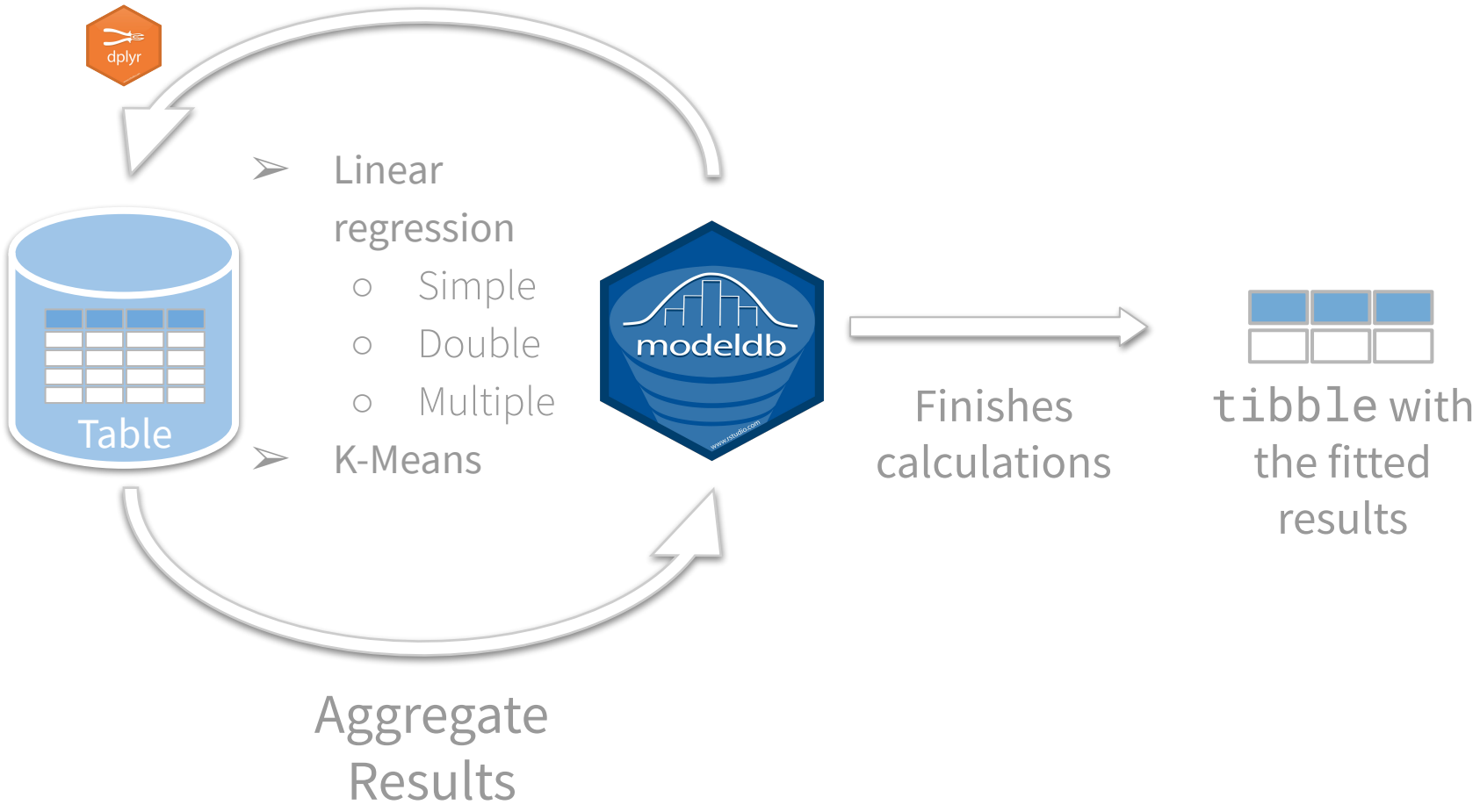
Today, we have parsers for 8 R models

1. Linear Regression - `lm()`
2. Generalized Linear - `glm()`
3. Random forest - `randomForest::randomForest()`
4. Ranger - `ranger::ranger()`
5. MARS - `earth::earth()`
6. Cubist - `Cubist::cubist()`
7. partykit tree - `partykit::ctree()`
8. XGBoost - `xgboost::xgb.Booster.complete()`

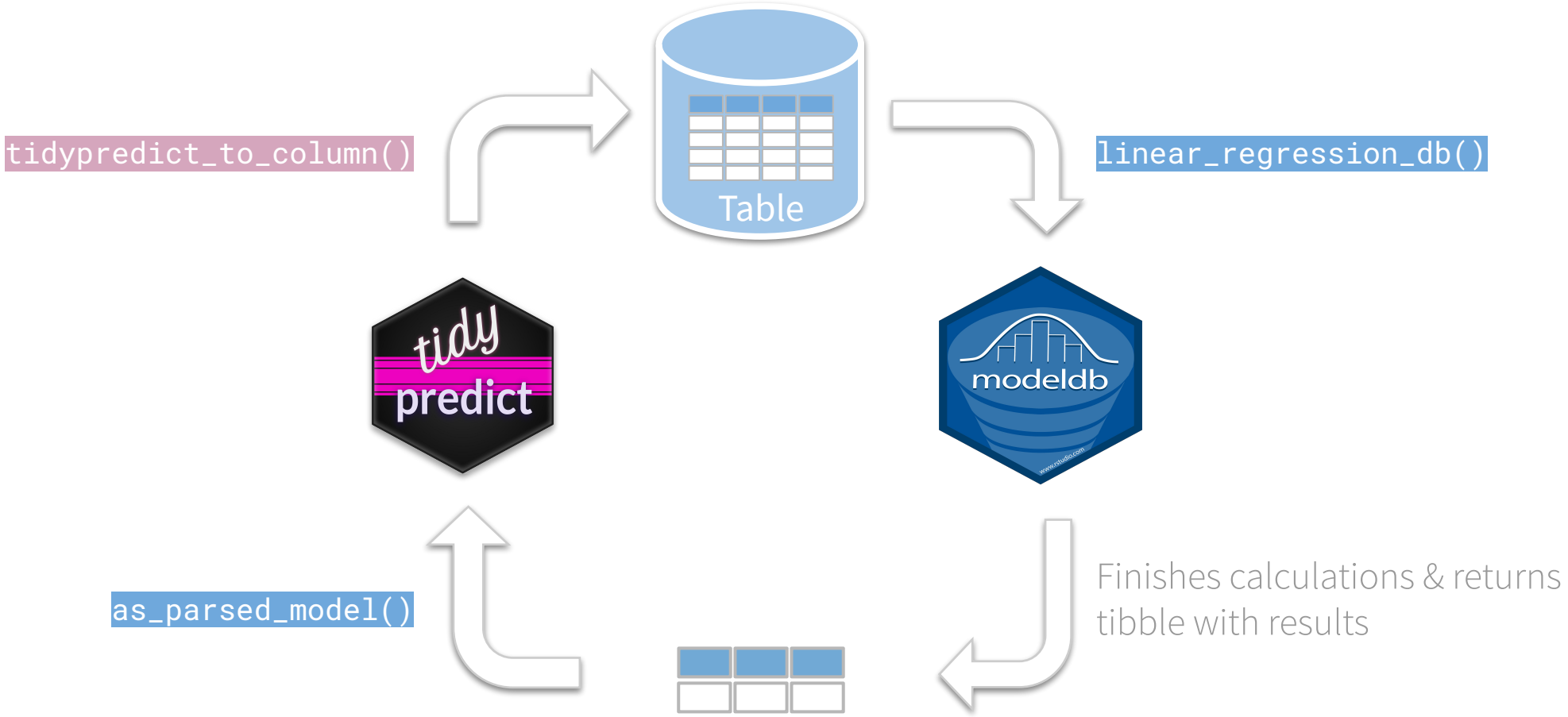


modeldb

`modeldb` estimates model in database



modeldb + tidypredict

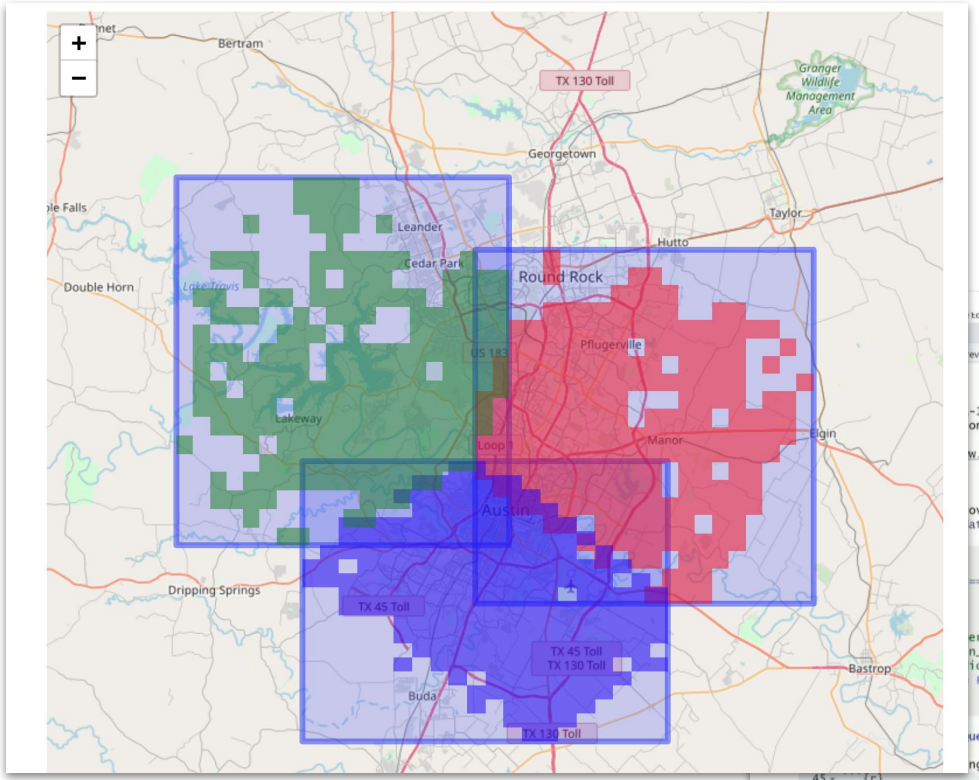


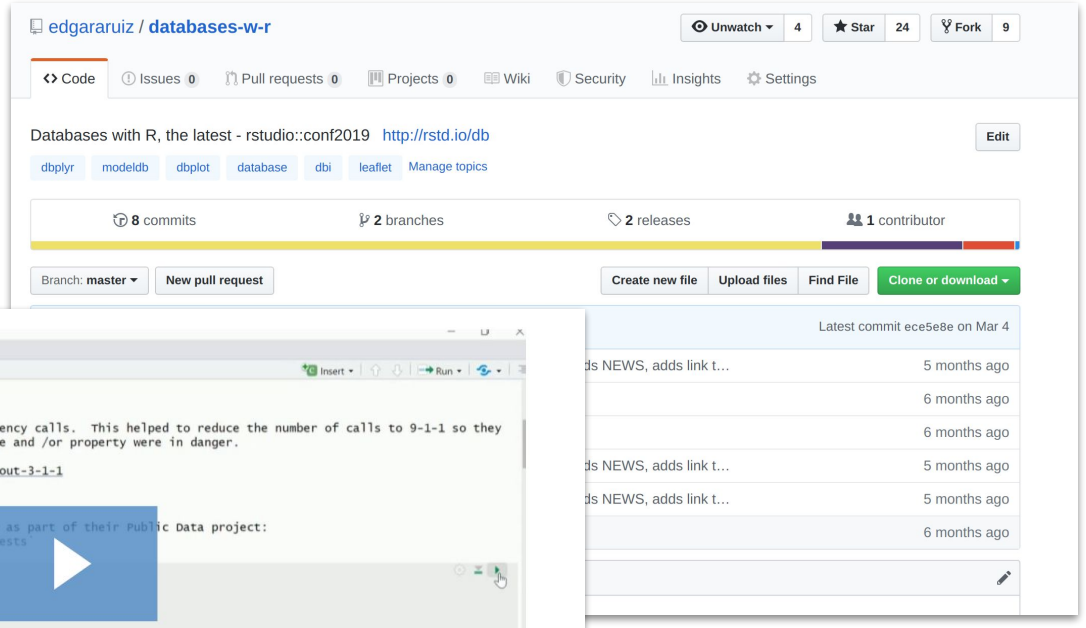
Demo

More...

K-means using `modeldb`

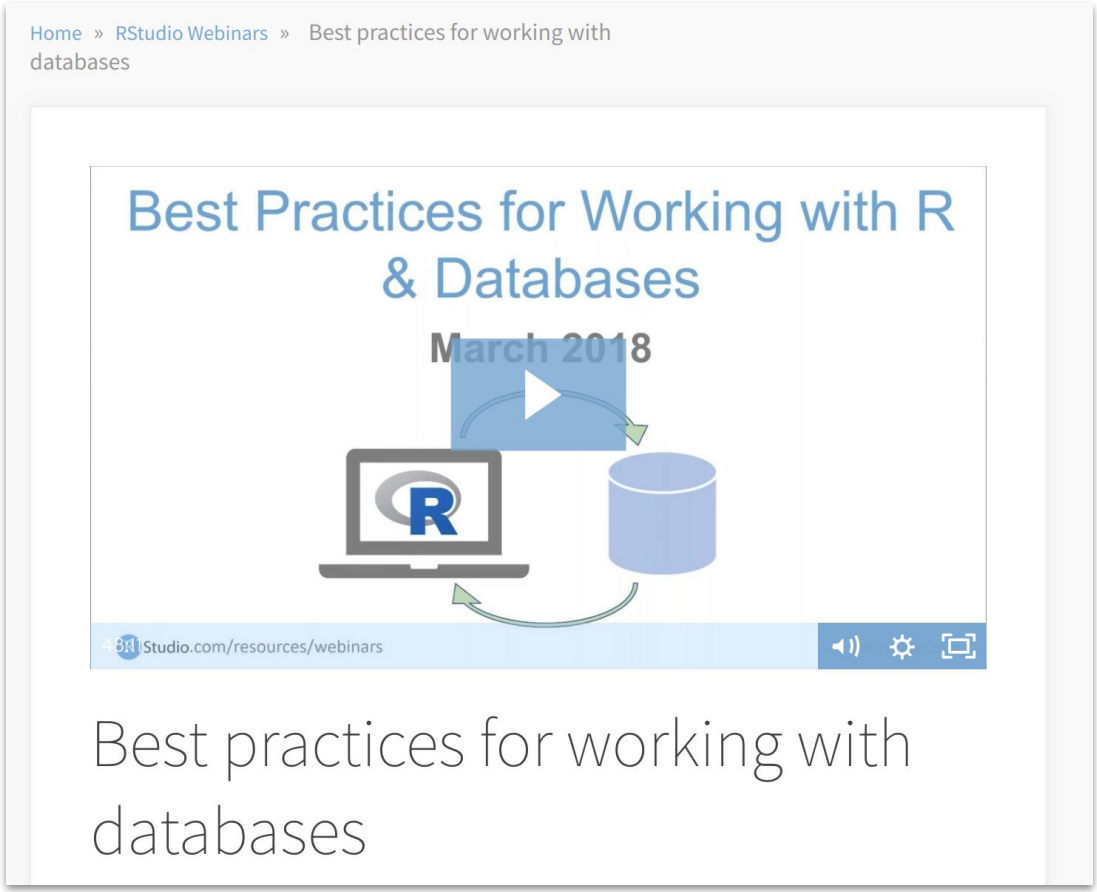
https://resources.rstudio.com/rstudio-conf-2019/databases-using-r-the-latest



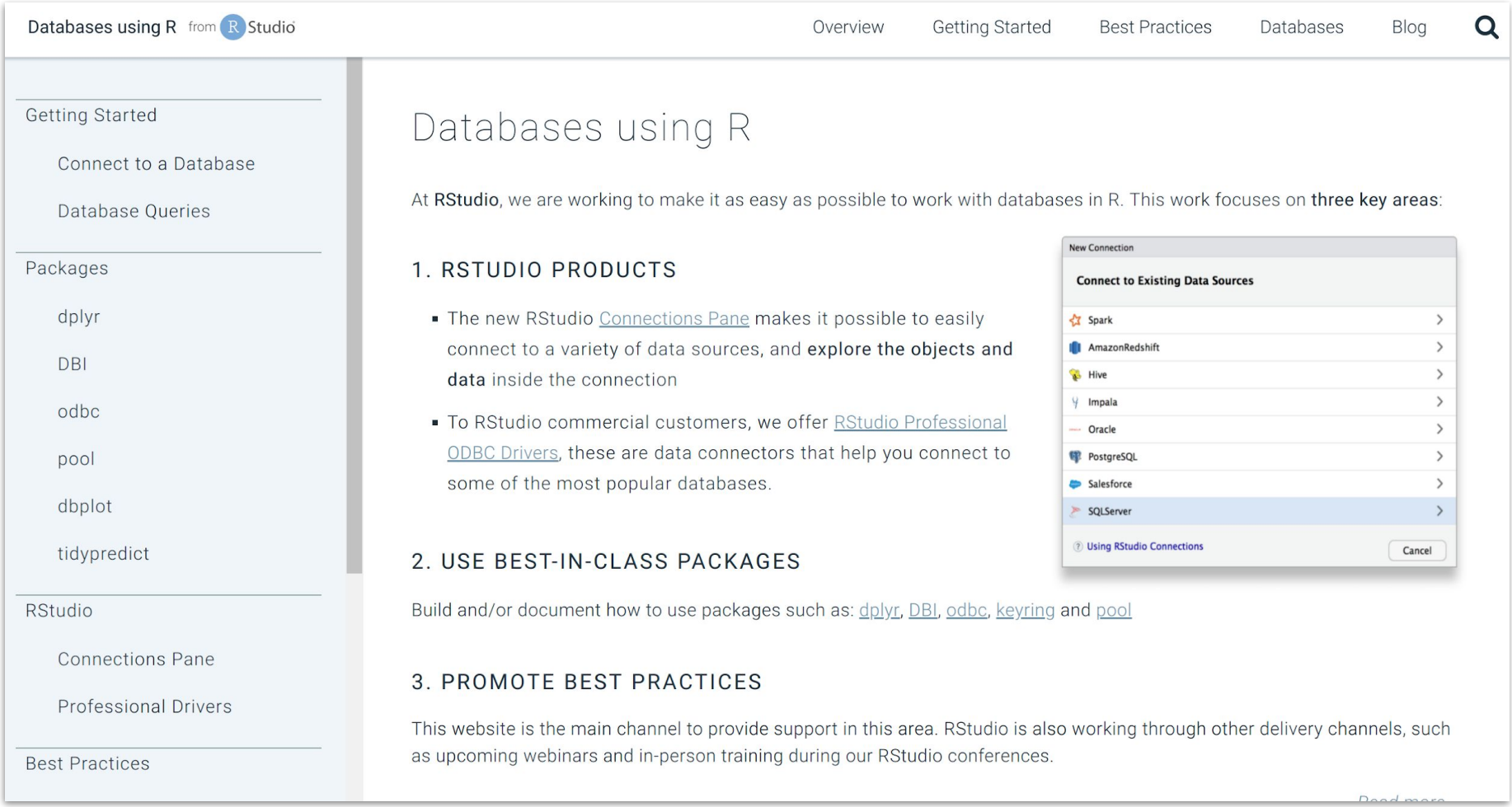


Intro material

https://resources.rstudio.com/webinars/best-practices-for-working-with-databases-march-2018-edgar-ruiz



db.rstudio.com



Thank you!

Photo by Gary Bendig on Unsplash

RStudio

db.rstudio.com