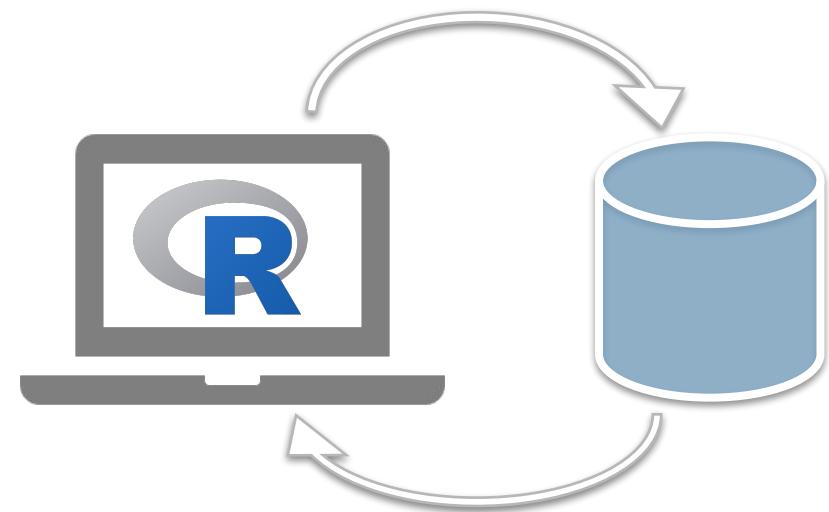


# Modeling in Databases with R

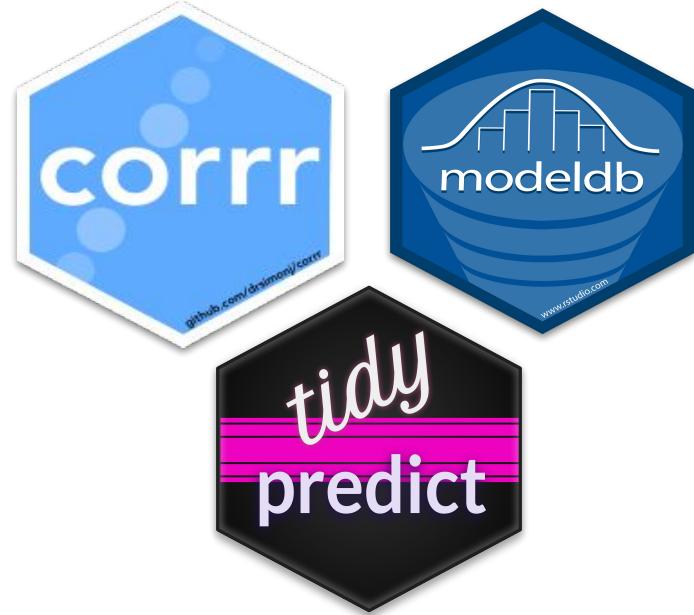
Edgar Ruiz

 theotheredgar

July 2019

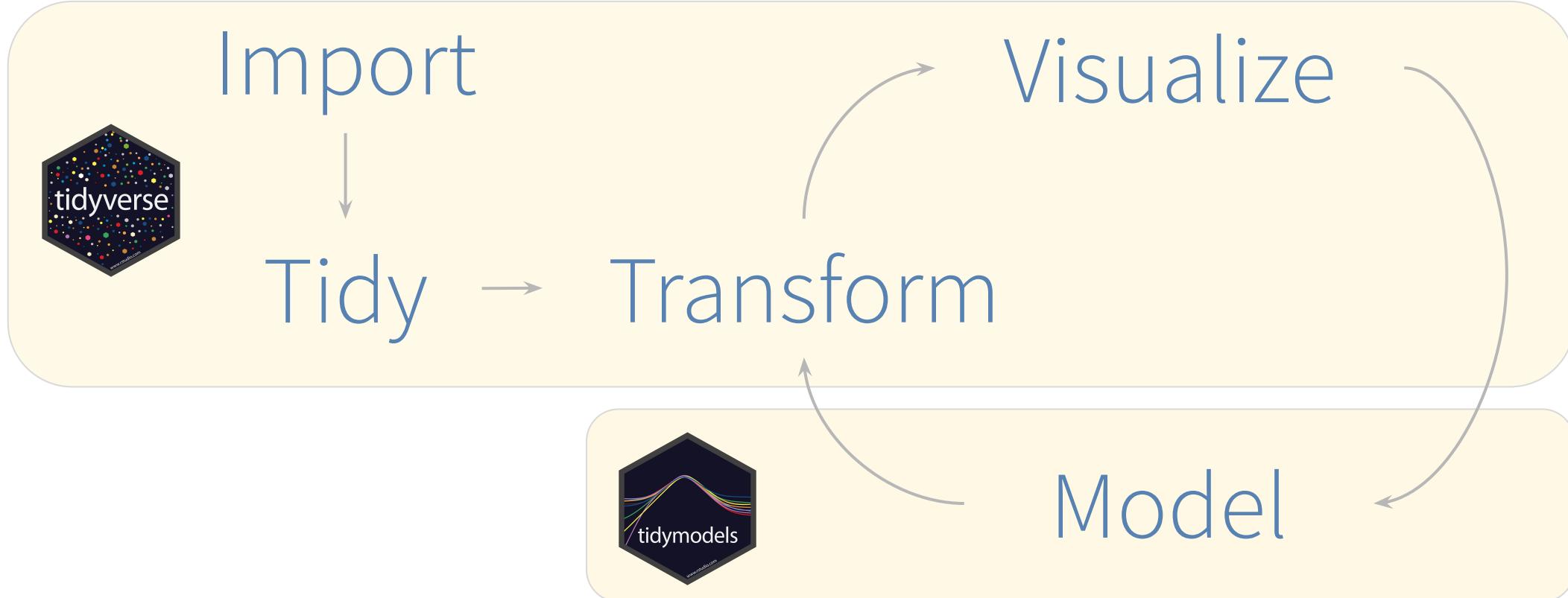


# Focus on three package...

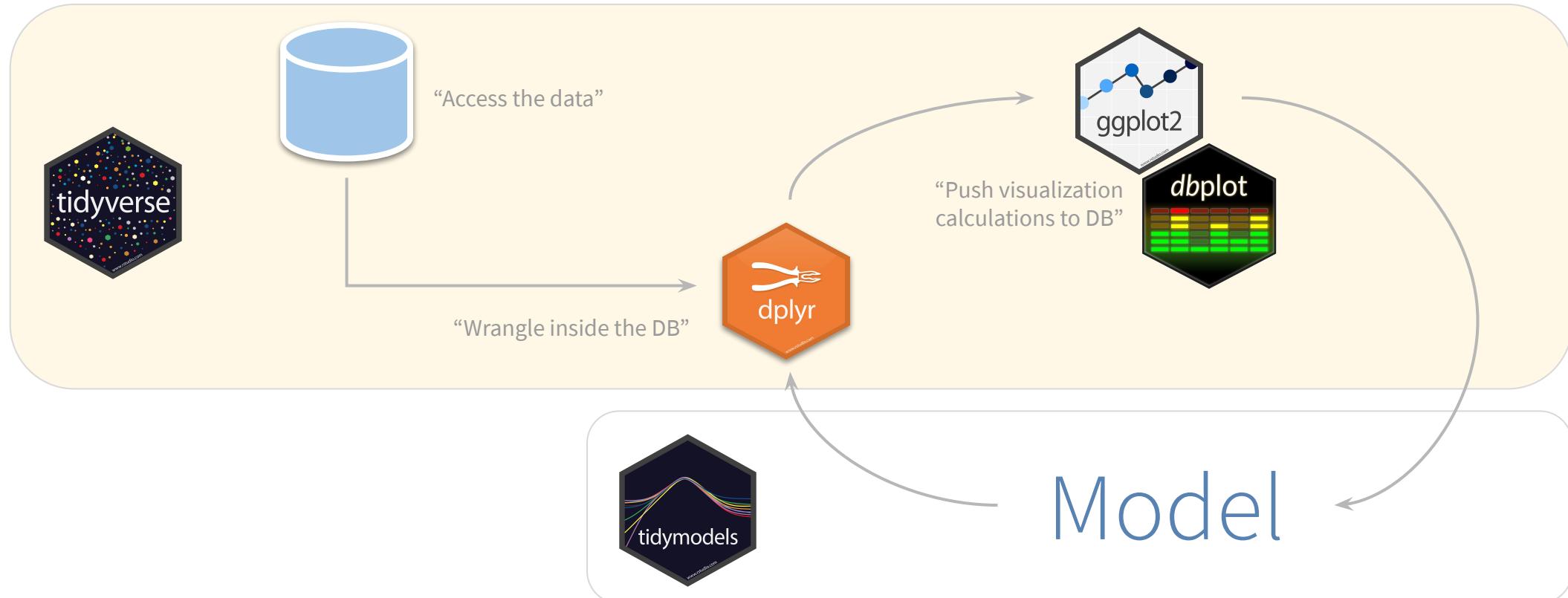


# Where do they fit?

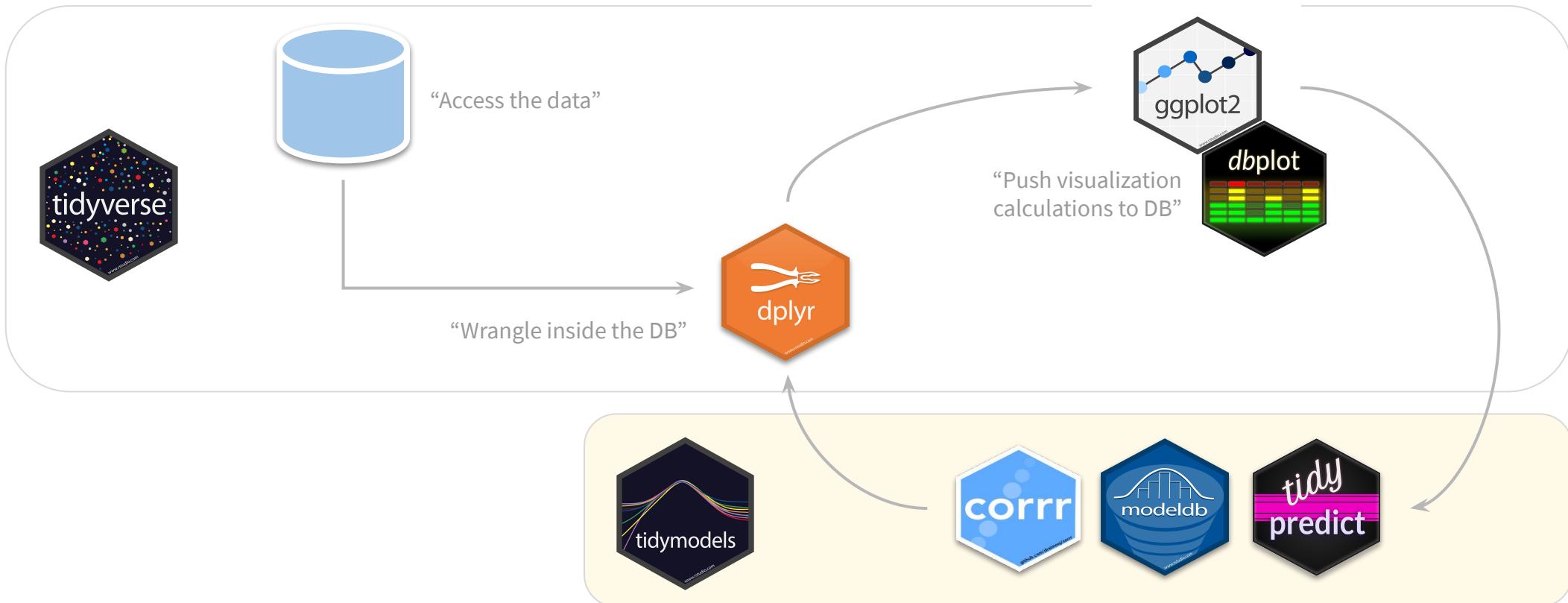
# Tidyverse



# Tidyverse + Databases



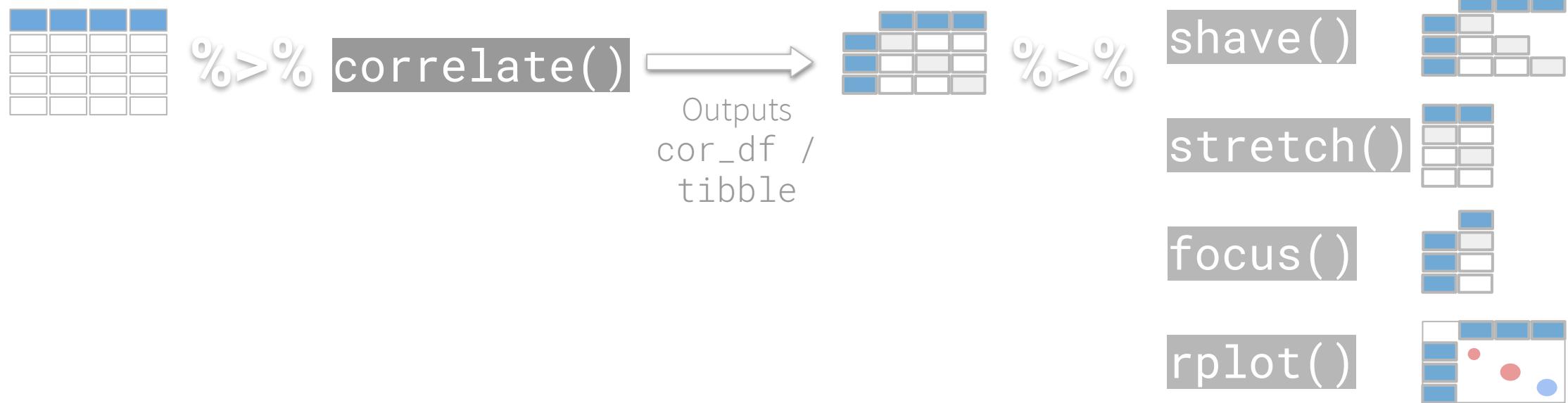
# Tidymodels + Databases



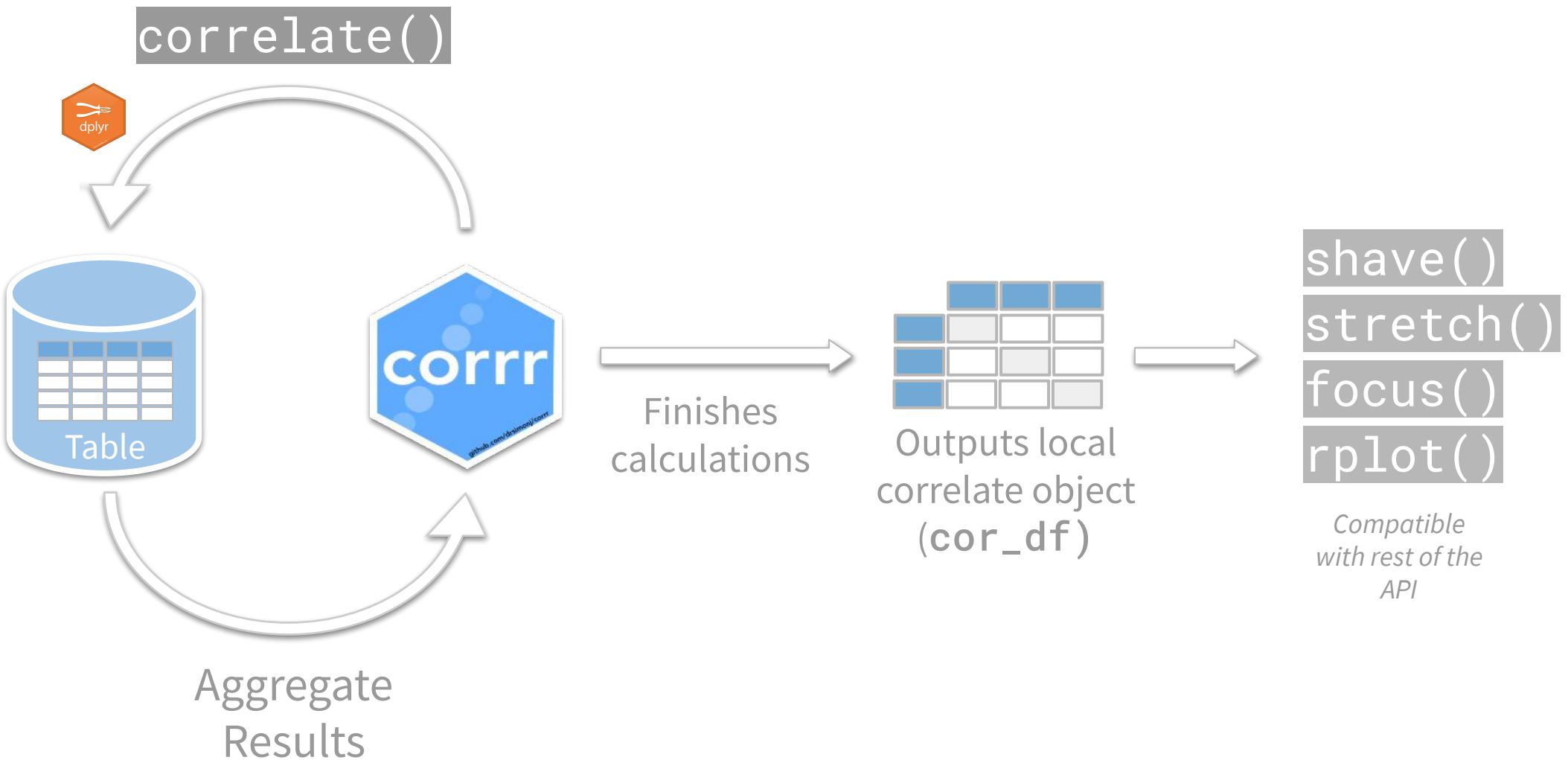
corr

# corrr package

corrr integrates with the tidyverse. It comes with specialized functions that make exploring even easier. Great for everyday analysis.

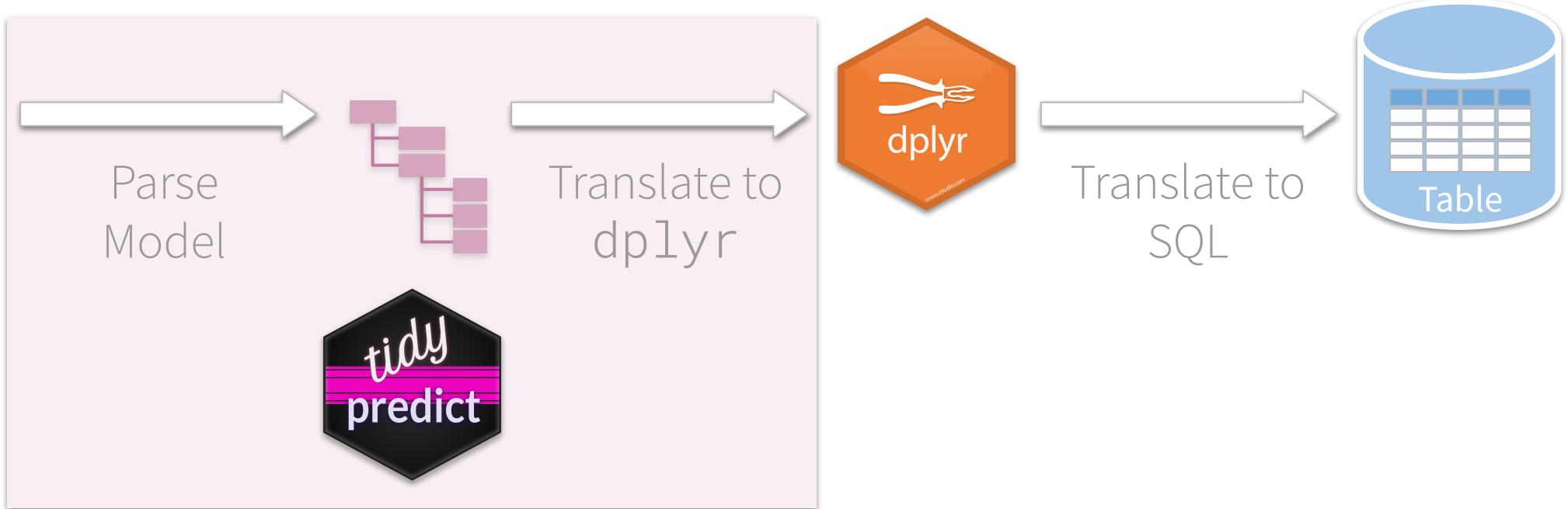
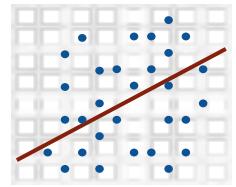


# corrr database integration

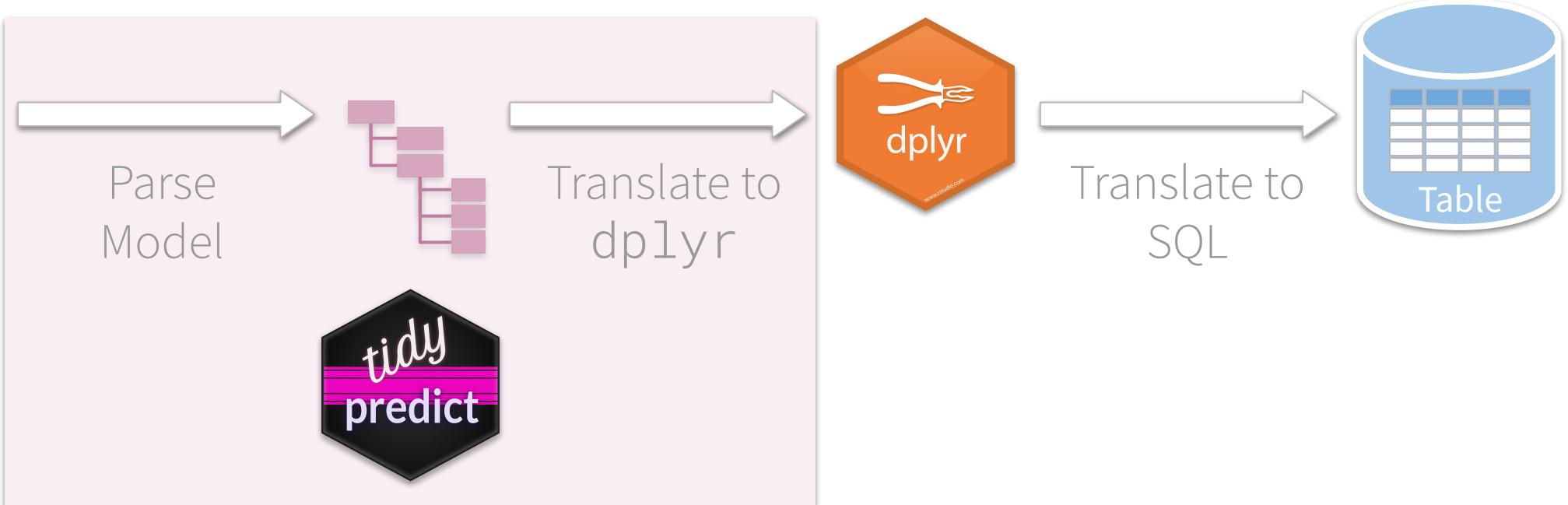
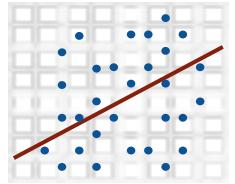


# tidypredict

# tidypredict workflow

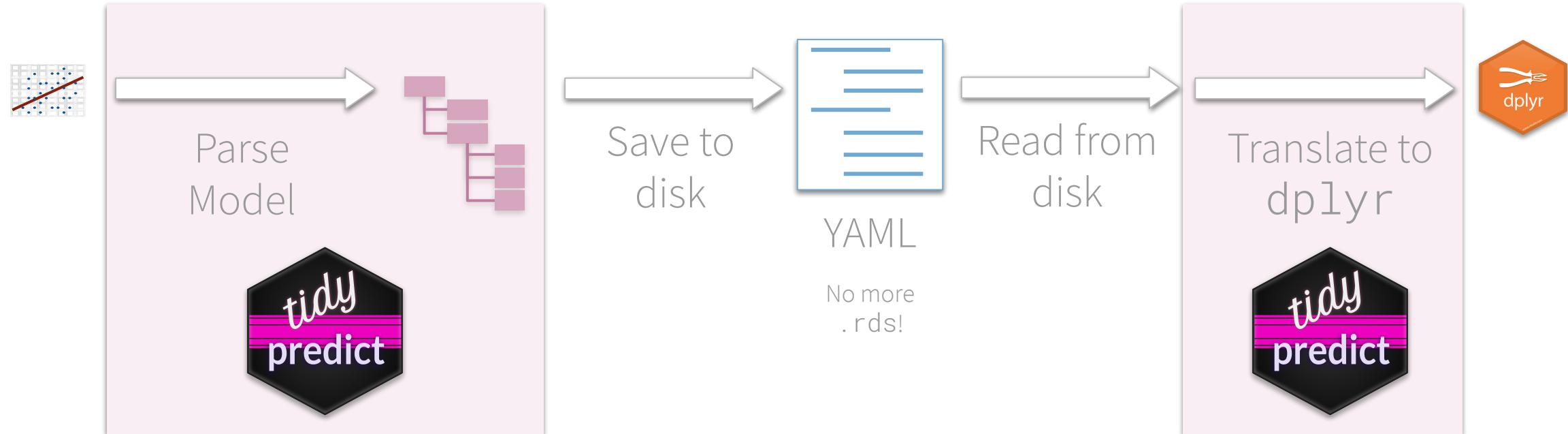


# tidypredict API



```
tidypredict_to_column()  
tidypredict_sql()  
tidypredict_fit()  
parse_model()
```

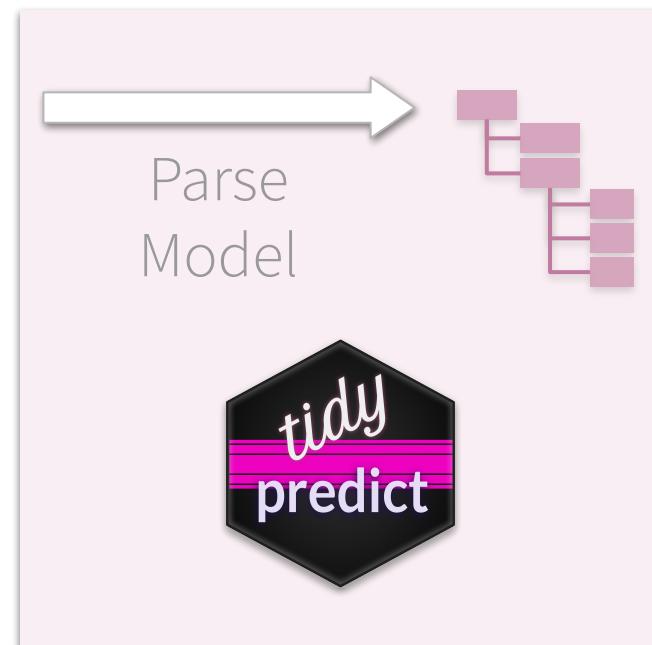
# Save the parsed model



# Today, we have parsers for 8 R models

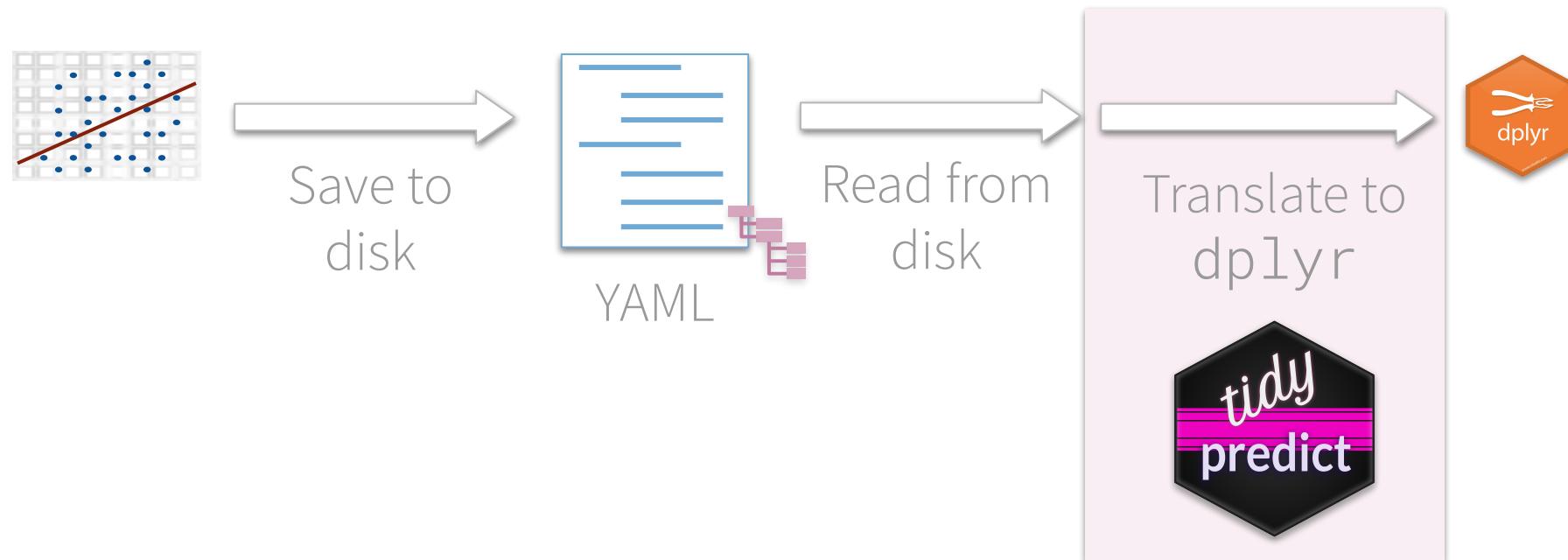


1. Linear Regression - `lm()`
2. Generalized Linear - `glm()`
3. Random forest - `randomForest::randomForest()`
4. Ranger - `ranger::ranger()`
5. MARS - `earth::earth()`
6. Cubist - `Cubist::cubist()`
7. partykit tree - `partykit::ctree()`
8. XGBoost - `xgboost::xgb.Booster.complete()`



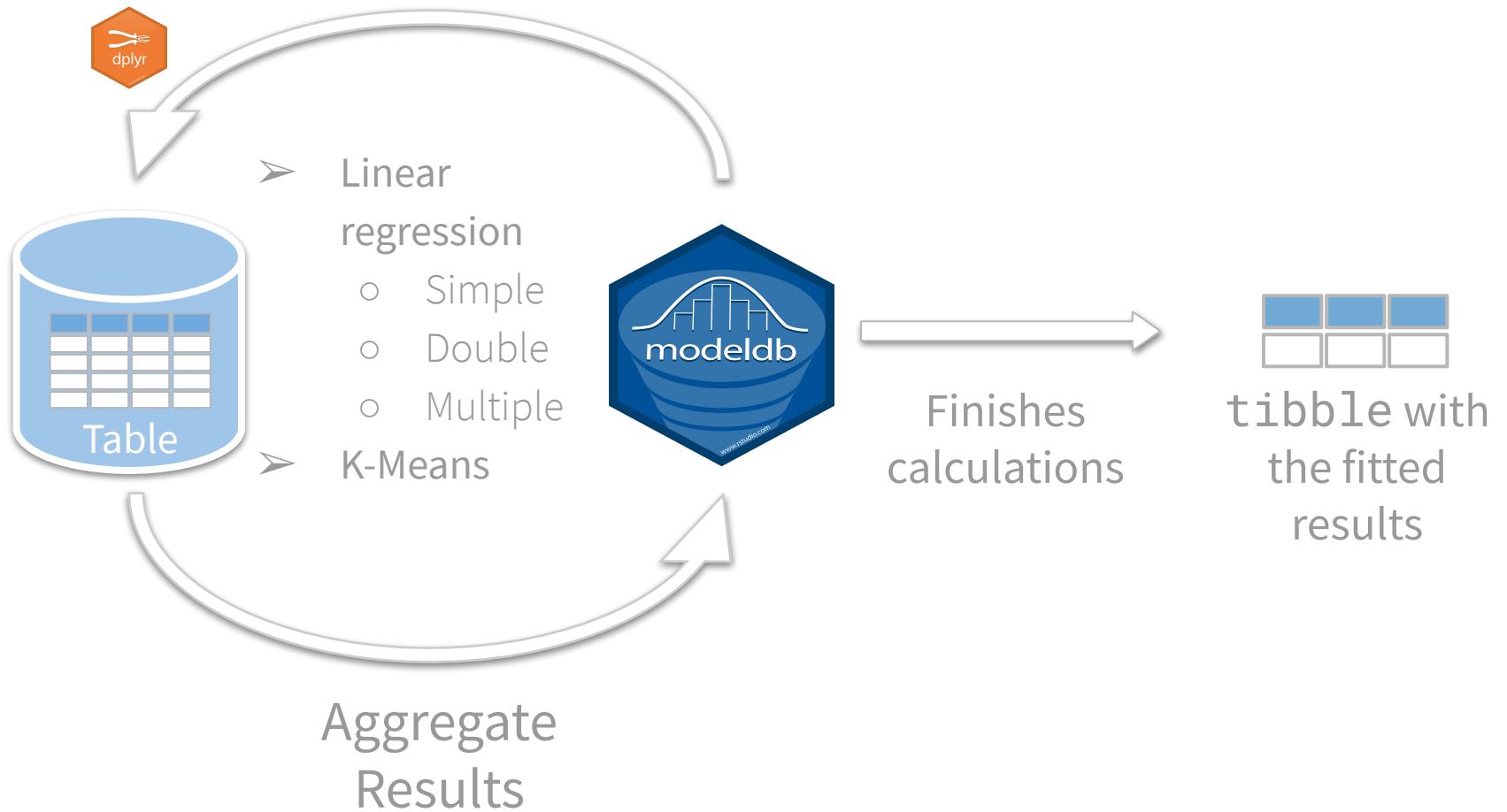
# Integrate models from other languages

`tidypredict` is able to use exported files from *non-R* scripts as long as they follow one of three parse model specs (regression, tree, xgb)

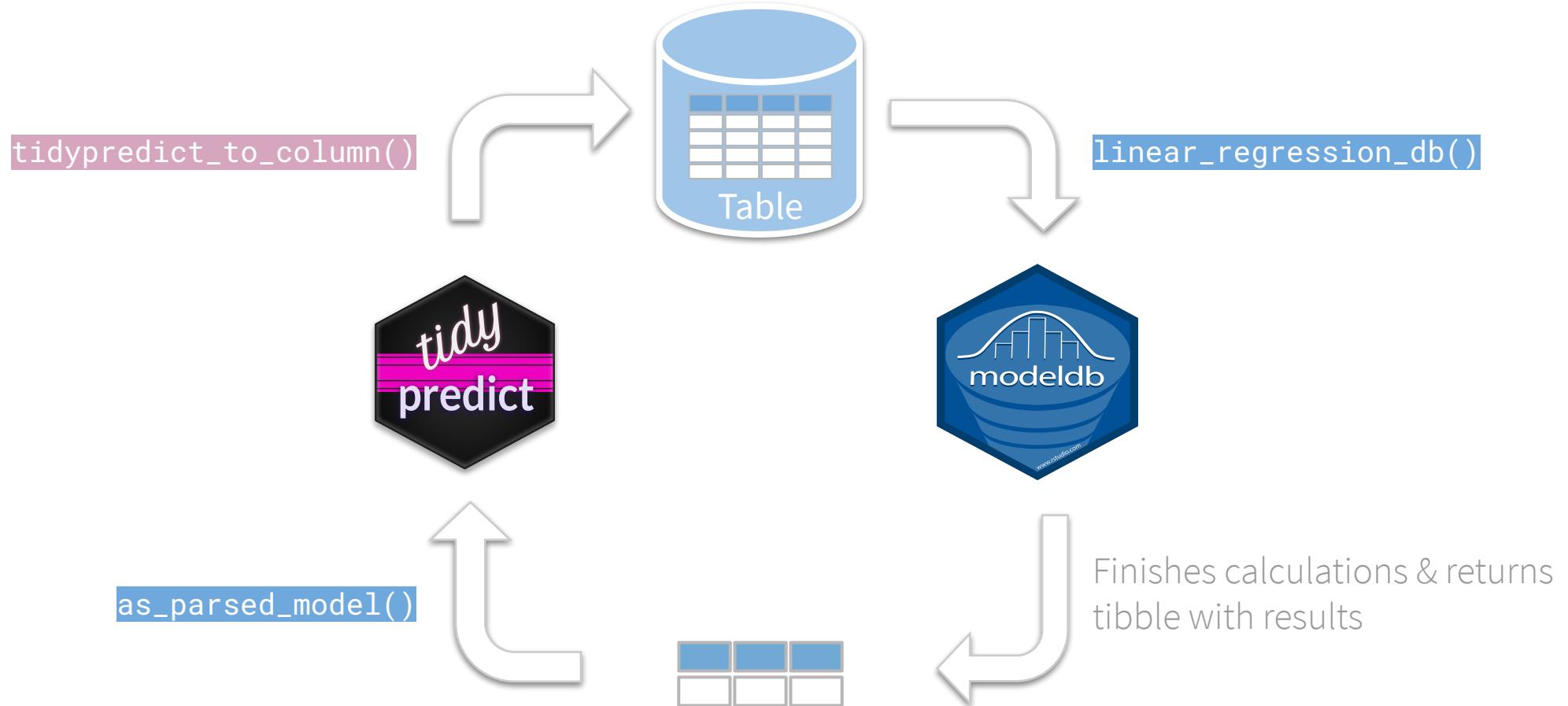


# modeldb

# modeldb runs models in database

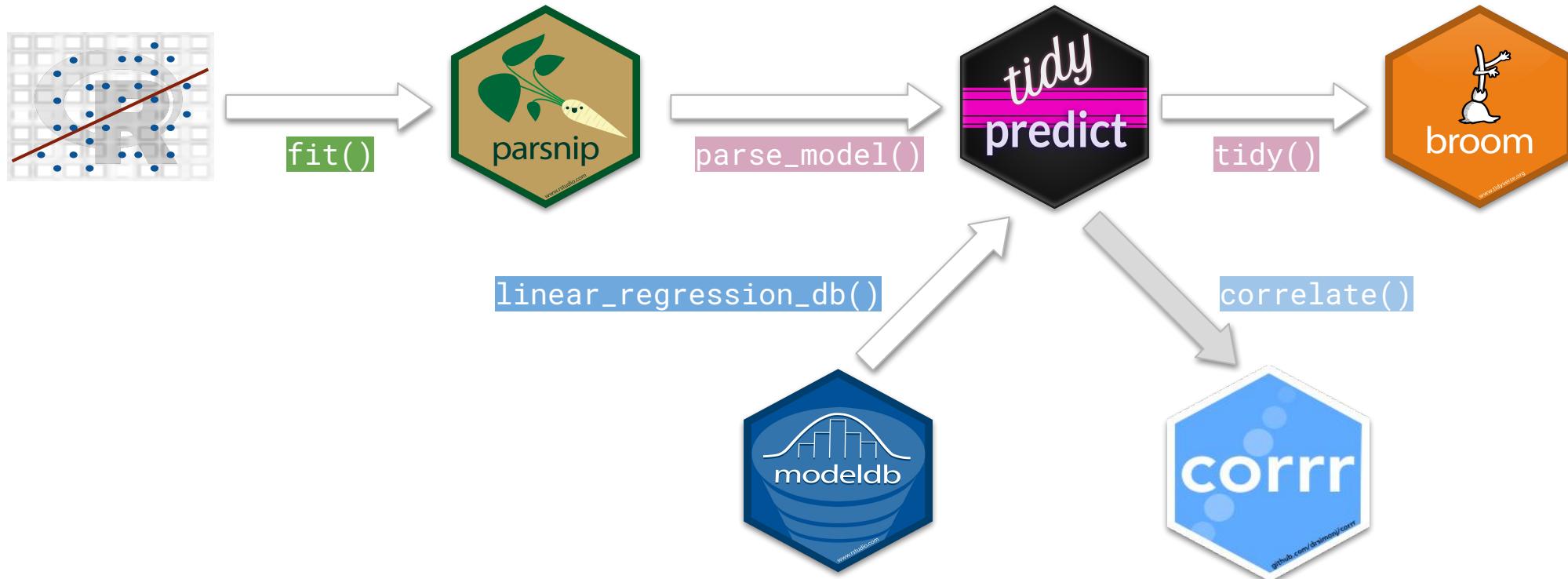


# modeldb + tidy predict

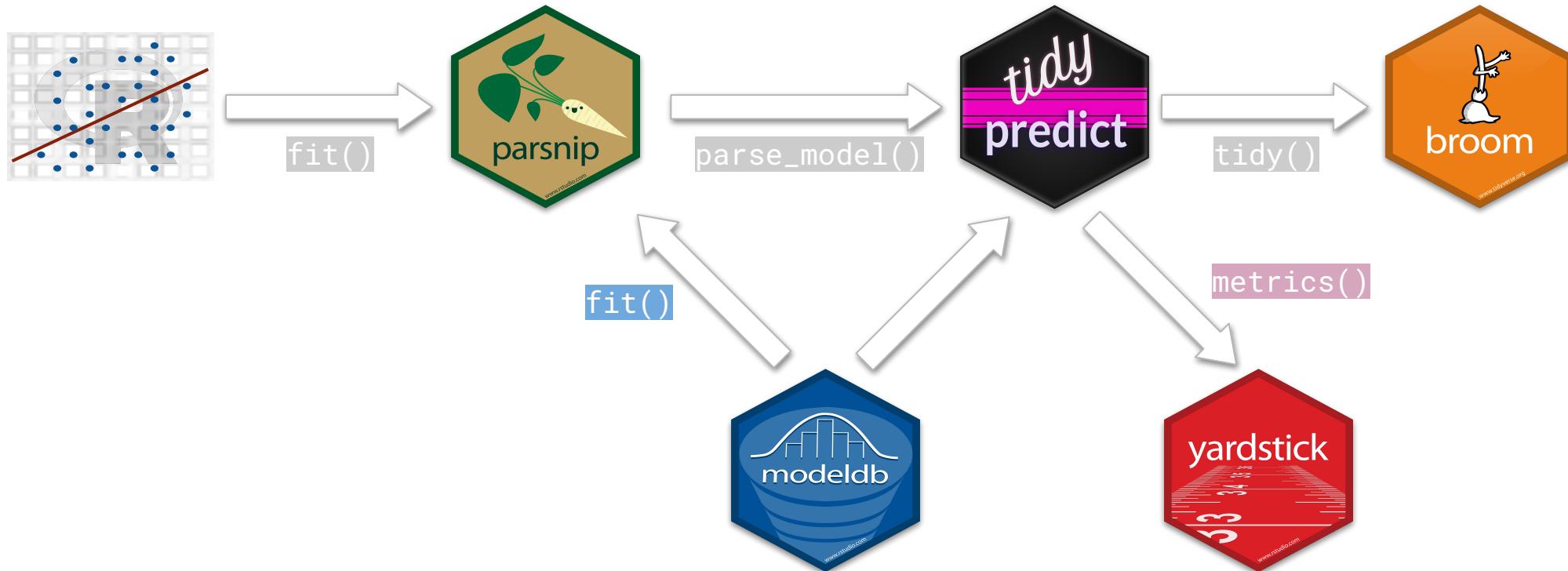


# Integration

# Today...



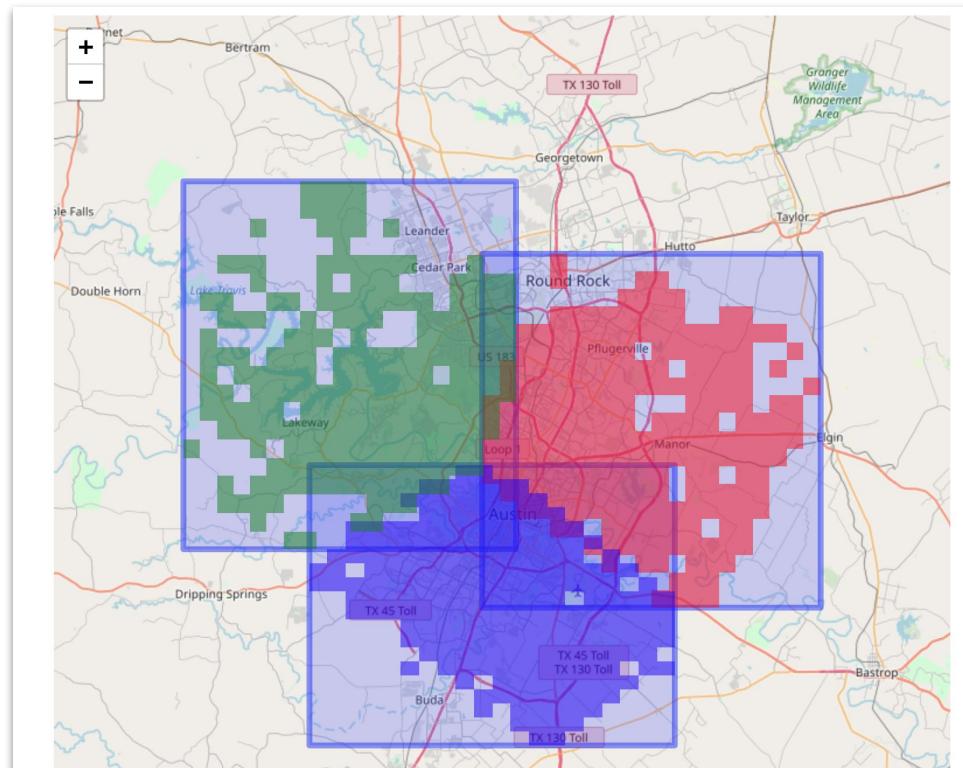
# Today...



More...

# K-means using modeldb

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



edgararuiz / databases-w-r

Code Issues Pull requests Projects Wiki Security Insights Settings

Databases with R, the latest - rstudio::conf2019 <http://rstudio.com/db>

dbplyr modeldb dbplot database dbi leaflet Manage topics

8 commits 2 branches 2 releases 1 contributor

Branch: master New pull request Create new file Upload files Find File Clone or download

Latest commit ece5e8e on Mar 4

ds NEWS, adds link t... 5 months ago

ds NEWS, adds link t... 6 months ago

ds NEWS, adds link t... 6 months ago

ds NEWS, adds link t... 5 months ago

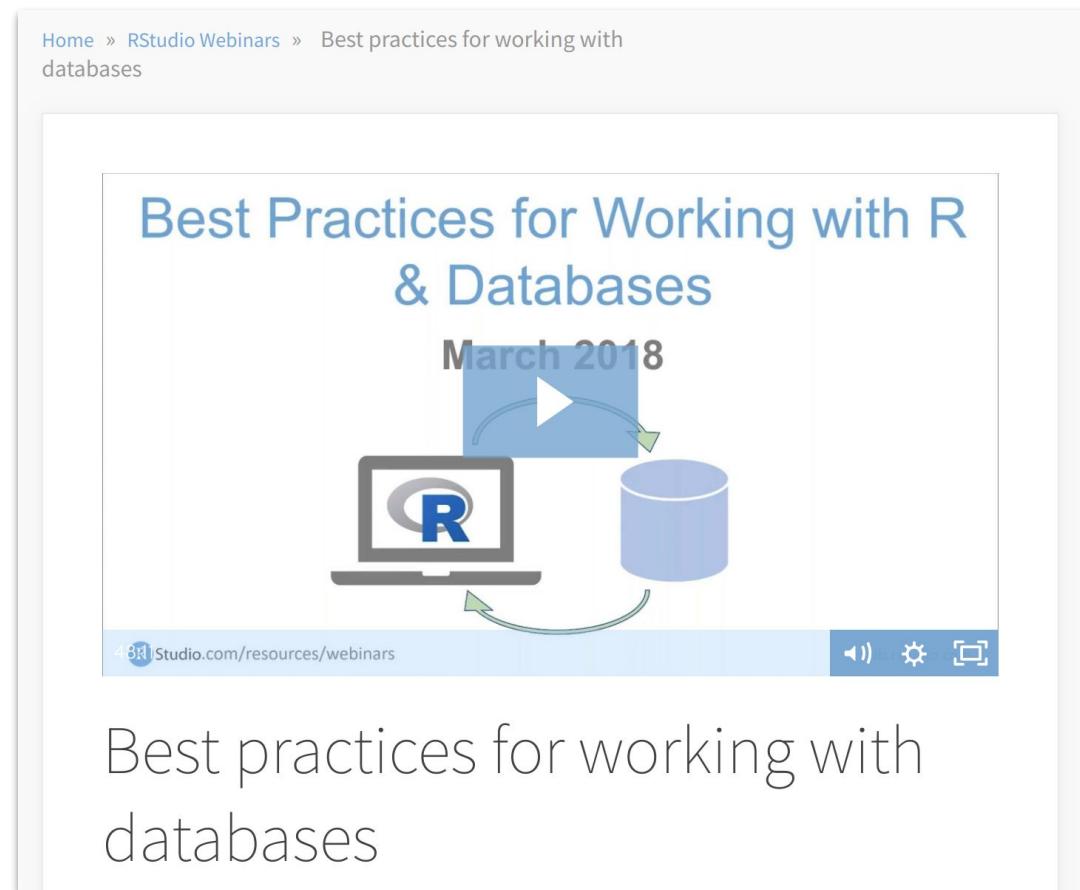
ds NEWS, adds link t... 5 months ago

ds NEWS, adds link t... 6 months ago

Databases using R: The latest -

# Intro material

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



Databases using R from  Studio

Overview Getting Started Best Practices Databases Blog 

---

Getting Started

- Connect to a Database
- Database Queries

---

Packages

- dplyr
- DBI
- odbc
- pool
- dbplot
- tidypredict

---

RStudio

- Connections Pane
- Professional Drivers

---

Best Practices

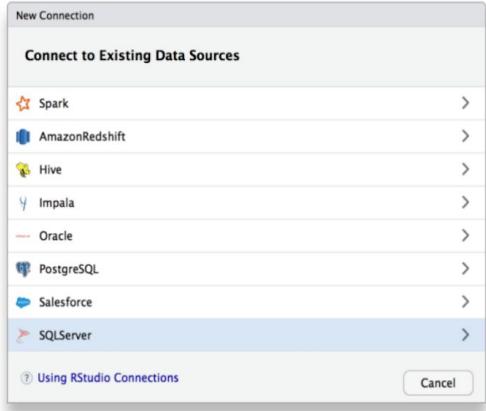
# Databases using R

At RStudio, we are working to make it as easy as possible to work with databases in R. This work focuses on **three key areas**:

- 1. RSTUDIO PRODUCTS**
  - The new RStudio [Connections Pane](#) makes it possible to easily connect to a variety of data sources, and [explore the objects and data](#) inside the connection
  - To RStudio commercial customers, we offer [RStudio Professional ODBC Drivers](#), these are data connectors that help you connect to some of the most popular databases.
- 2. USE BEST-IN-CLASS PACKAGES**

Build and/or document how to use packages such as: [dplyr](#), [DBI](#), [odbc](#), [keyring](#) and [pool](#)
- 3. PROMOTE BEST PRACTICES**

This website is the main channel to provide support in this area. RStudio is also working through other delivery channels, such as upcoming webinars and in-person training during our RStudio conferences.



New Connection

Connect to Existing Data Sources

- Spark
- AmazonRedshift
- Hive
- Impala
- Oracle
- PostgreSQL
- Salesforce
- SQLServer

Using RStudio Connections Cancel

# Thank you!



Photo by [Gary Bendig](#) on [Unsplash](#)