

Databases with RStudio

Creating Shiny dashboards performant at scale



Using Databases with R

- You can use R with almost any data source



- But historically it's not always easy

Obstacles when using databases with R

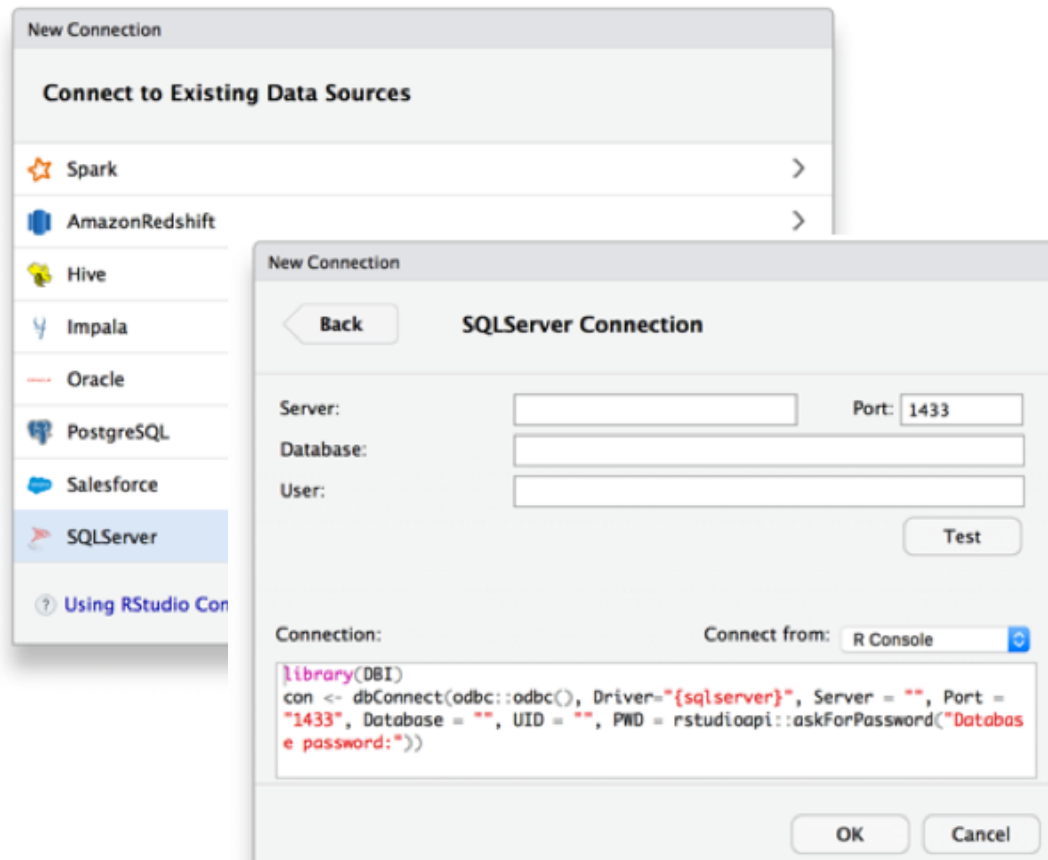
Obstacles	Examples
Hard to establish connections and access data	R users often use one tool for SQL and another tool for R
No consistent toolset or language	R users interchange SQL and R to talk to the exact same data
Setup and configuration is confusing	R users unable to access their databases
No centralized place to get information	R users get their information from package manuals

What's new from RStudio

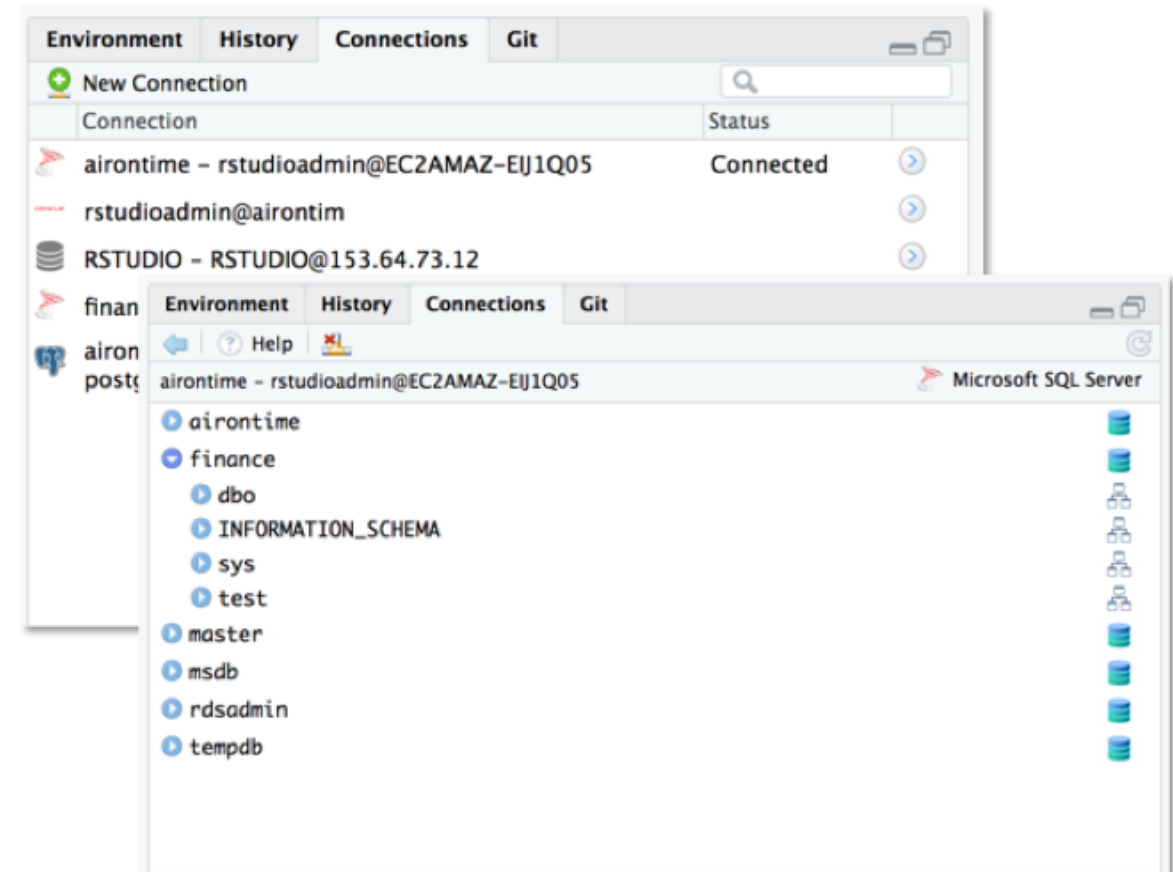


RStudio version 1.1 new features

Connection wizard

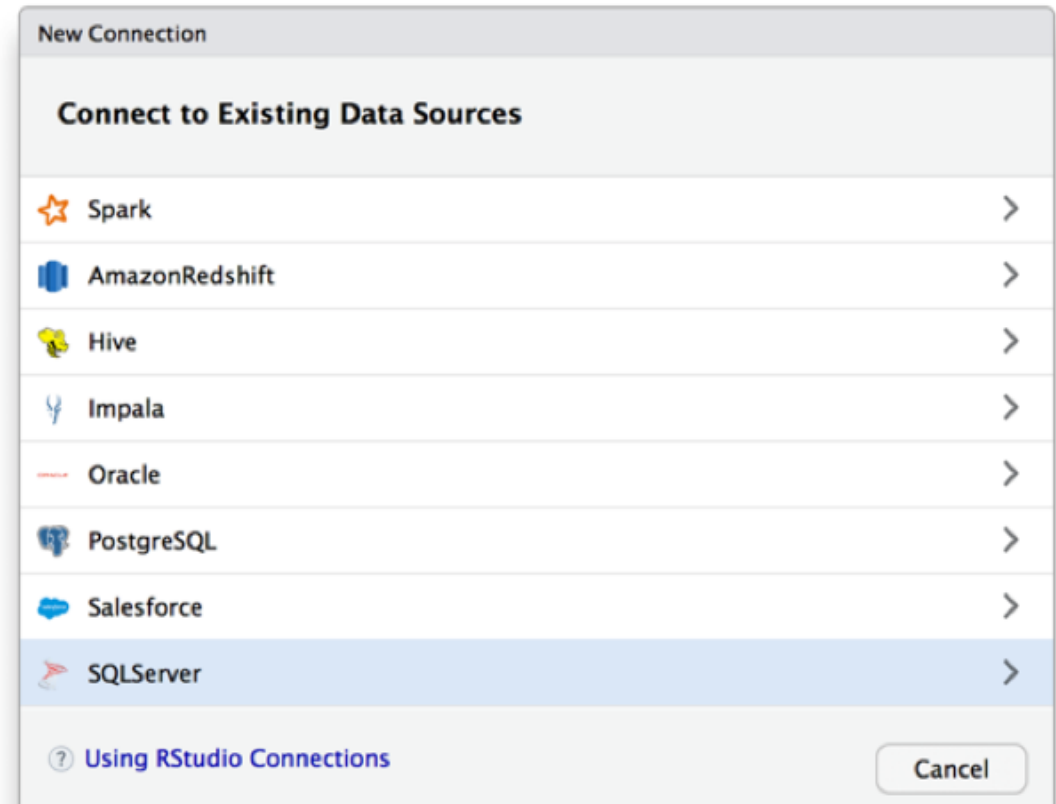


Connections tab



Connection methods

1. ODBC drivers
2. Data source names (DSN)
3. R Snippets
4. R Packages



Improved R Packages

- DBI
 - Standard database interface for R
 - Query your database with SQL
 - Do database operations
- dplyr
 - Query your database with R code
 - Generalized SQL backend
 - dbplyr [New!] translates dplyr syntax to SQL for specific databases

dplyr syntax

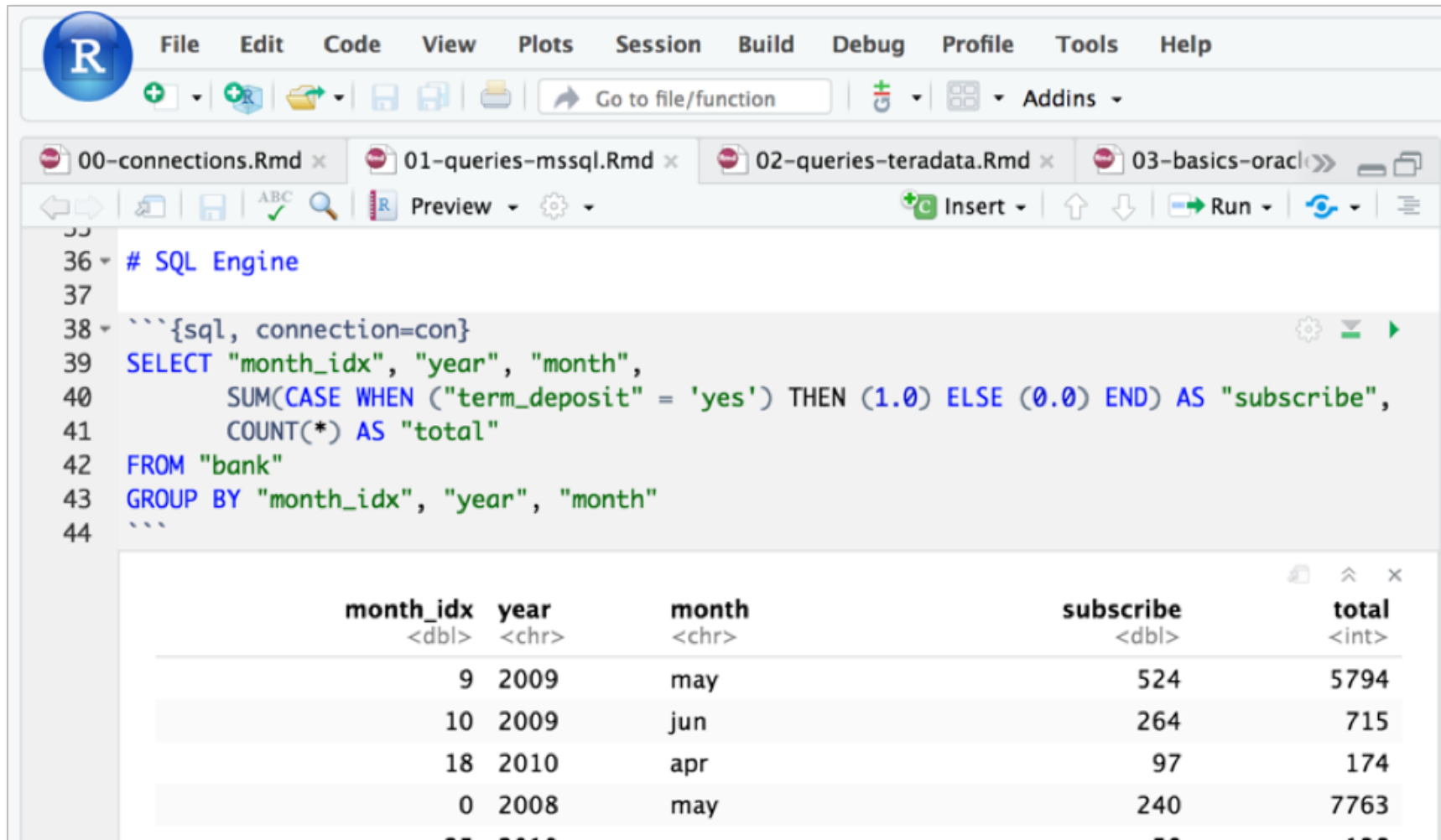
Use dplyr verbs

```
flights %>%  
  group_by(name) %>%  
  tally()
```

Write SQL statements

```
SELECT "name",  
COUNT(*) AS "n"  
FROM "vwFlights"  
GROUP BY "name"
```


SQL Language Engine for: *R Markdown and R Notebooks*



The screenshot shows the R Studio interface with a menu bar (File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help) and a toolbar. The active document is '03-basics-oracle.Rmd'. The code editor displays an SQL query within a code chunk. The query selects data from a table named 'bank', grouped by month_idx, year, and month. It includes a conditional sum for 'term_deposit' = 'yes' and a count of rows. The results are displayed in a table below the code.

```
36 # SQL Engine
37
38 ```{sql, connection=con}
39 SELECT "month_idx", "year", "month",
40        SUM(CASE WHEN ("term_deposit" = 'yes') THEN (1.0) ELSE (0.0) END) AS "subscribe",
41        COUNT(*) AS "total"
42 FROM "bank"
43 GROUP BY "month_idx", "year", "month"
44 ```
```

month_idx <dbl>	year <chr>	month <chr>	subscribe <dbl>	total <int>
9	2009	may	524	5794
10	2009	jun	264	715
18	2010	apr	97	174
0	2008	may	240	7763

New R package for connecting with ODBC

odbc [New!]

- Connects R to any data source via ODBC
- Bring your own driver or use RStudio Professional Drivers
- DBI compliant
- Actively developed
- Designed for performance

Database connection methods

odbc [New!]

- Connects R to any data source via ODBC
- Bring your own driver or use RStudio Professional Drivers
- DBI compliant
- Actively developed
- Designed for performance

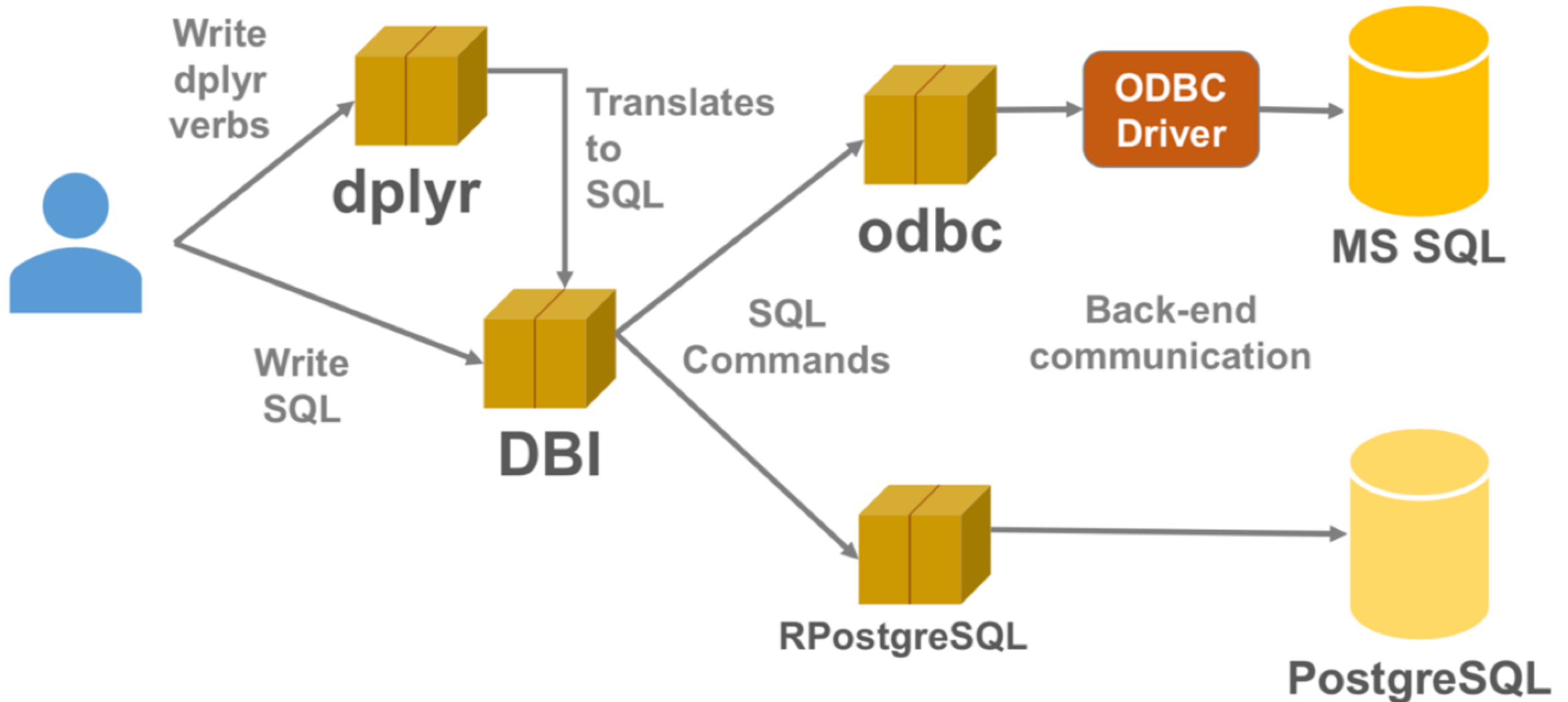
R Packages

- RSQLite
- rpostgres
- bigrquery
- [...]

RJDBC

- Connect to any data source via JDBC
- Requires JDBC driver
- Requires Java

Establish a database connection with RStudio

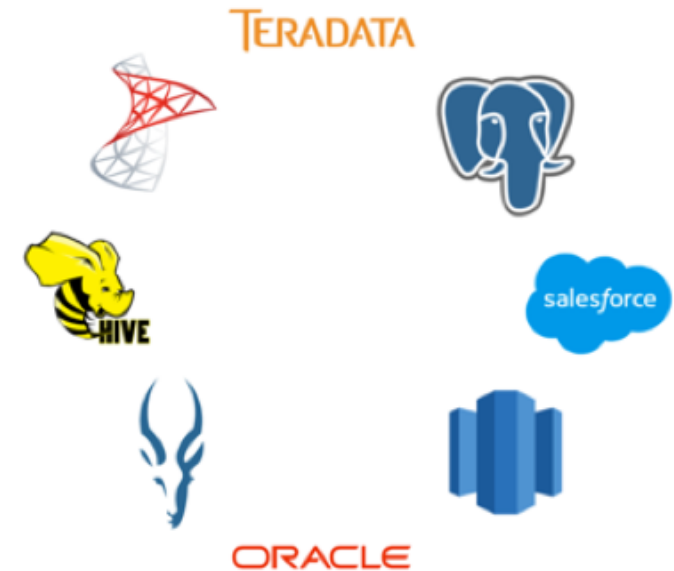


RStudio Professional Drivers



RStudio Professional Drivers

- Coverage
 - Connect to many popular databases
- Convenience
 - Easy to install and configure
- Consistent
 - Use with all RStudio pro products
- Professional
 - Supported for production level work



Teradata

PostgreSQL

Salesforce

Redshift

Oracle

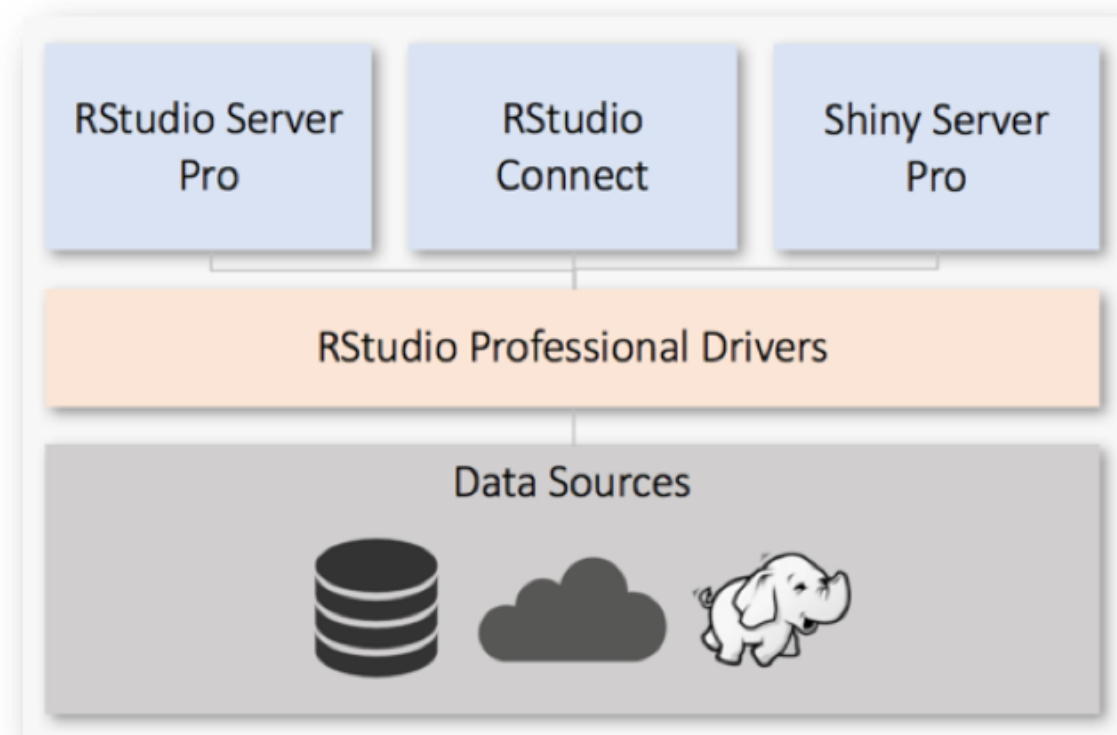
Impala

Hive

SQL Server

RStudio Professional Drivers in Production

- Using R and databases in production
 - Shiny applications
 - R Markdown reports
- Benefits of pro drivers
 - Support for the entire workflow
 - Professional license
 - Consistent dev/test/prod environment

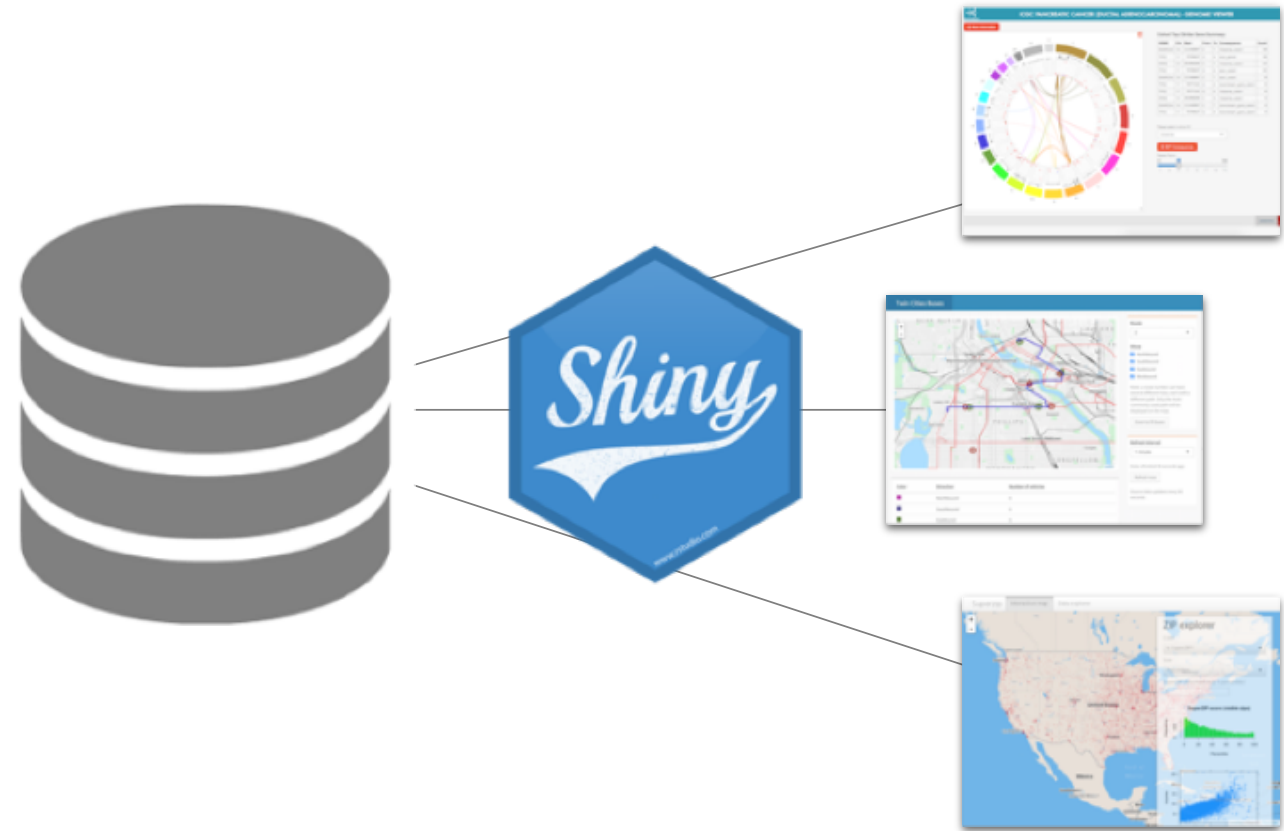


Databases with Shiny



Shiny applications and databases

- Develop and deploy Shiny applications that depend on databases
- Use RStudio professional products when hosting Shiny in production systems



Push computation to the database



Database Demos

- Using R with Databases
 - A deep dive into connections and dplyr
- Using Shiny with Databases
 - Simple Shiny app
 - Extracting data into a dashboard
 - Pushing compute into a dashboard
- Scoring data
 - Using API's
 - Using R and SQL
- <https://github.com/nwstephens/databases-and-shiny>

Learn more at db.rstudio.com

The screenshot shows the 'Databases using R' page from RStudio. The page has a navigation bar at the top with links to 'dplyr', 'DBI', 'Best Practices', 'Databases', and 'Advanced'. On the left side, there is a sidebar with a 'Packages' section containing links to 'dplyr', 'DBI', 'odbc', and 'pool', and an 'RStudio' section containing a link to 'Connections Pane'. Below the sidebar, there is a 'Best Practices' section with links to 'Setting up ODBC Drivers', 'Run Queries Safely', 'Securing Deployed Content', 'Securing Credentials', 'Making Scripts Portable', and 'Creating Visualizations'. The main content area is titled 'Databases using R' and contains an introductory paragraph followed by three bullet points: 'Developing best-in-class packages', 'Integrating with RStudio', and 'Promoting best practices'. Below this, there is a section titled 'Explore databases inside RStudio v1.1' which includes a screenshot of the RStudio interface showing the 'Connections' pane and a list of databases. Callout boxes highlight features like 'View all open DB connections', 'New Connections pane', 'Navigate databases, schemas, tables & fields', and 'Click to preview the connection'. To the right of the screenshot, there is a paragraph explaining the new RStudio Connections Pane.

Databases using R from RStudio

dplyr DBI Best Practices Databases Advanced

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:

- **Developing best-in-class packages** These packages include: `dplyr`, `odbc`, `keyring`, `pool` and others.
- **Integrating with RStudio** Starting with version 1.1, RStudio helps you write the R code you need to connect to a database, remembers connections you've made, and allows you to explore the objects and data in your database.
- **Promoting best practices** This website is the main way we provide support in this area. RStudio is also working through other delivery channels, like upcoming webinars and in-person training at [rstudio::conf 2018](#)

Explore databases inside RStudio v1.1

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.