

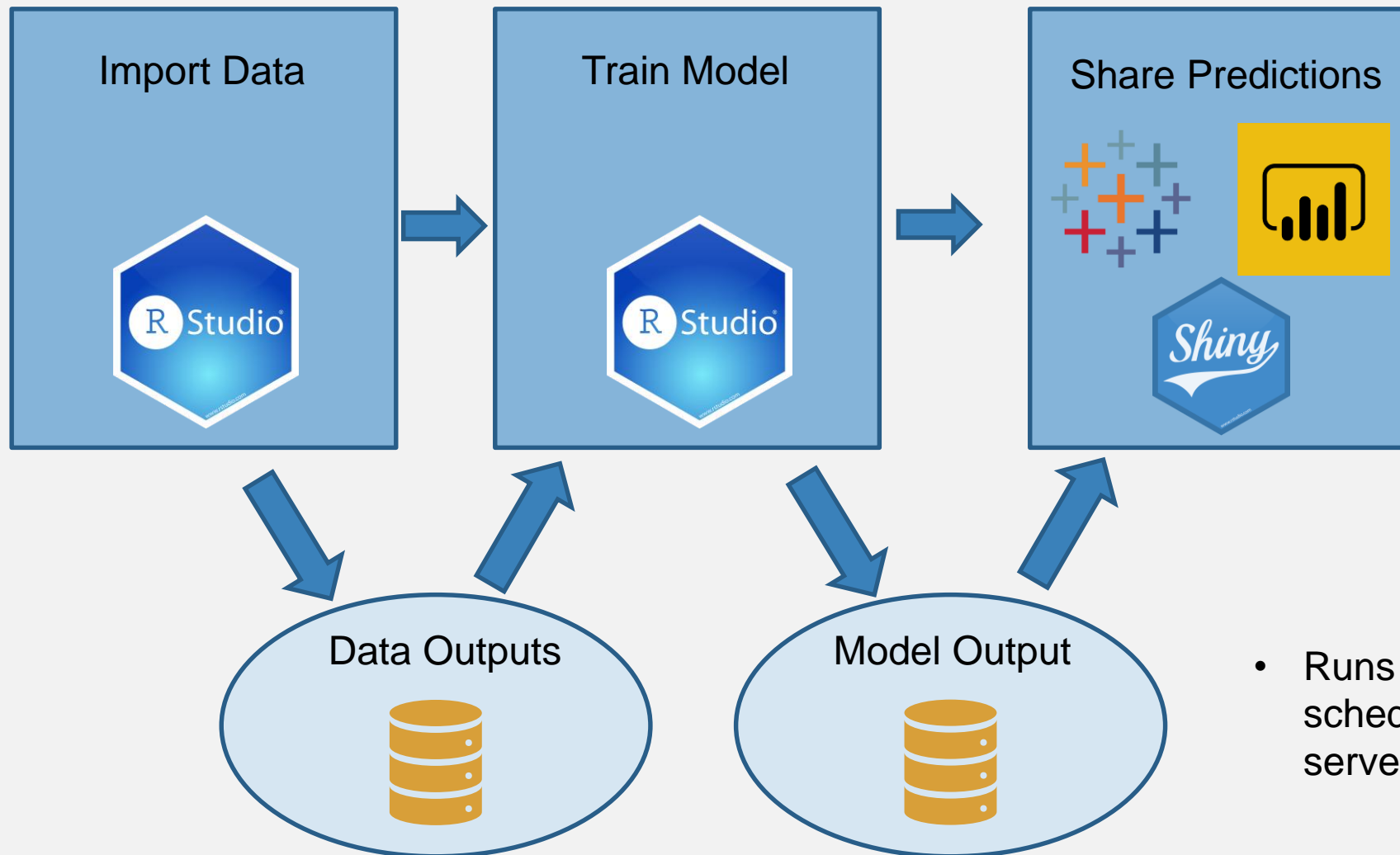
Deploying End-to-End Data Science Production Workflows in R

Alyssa Toay

Lead Data Scientist at General Mills



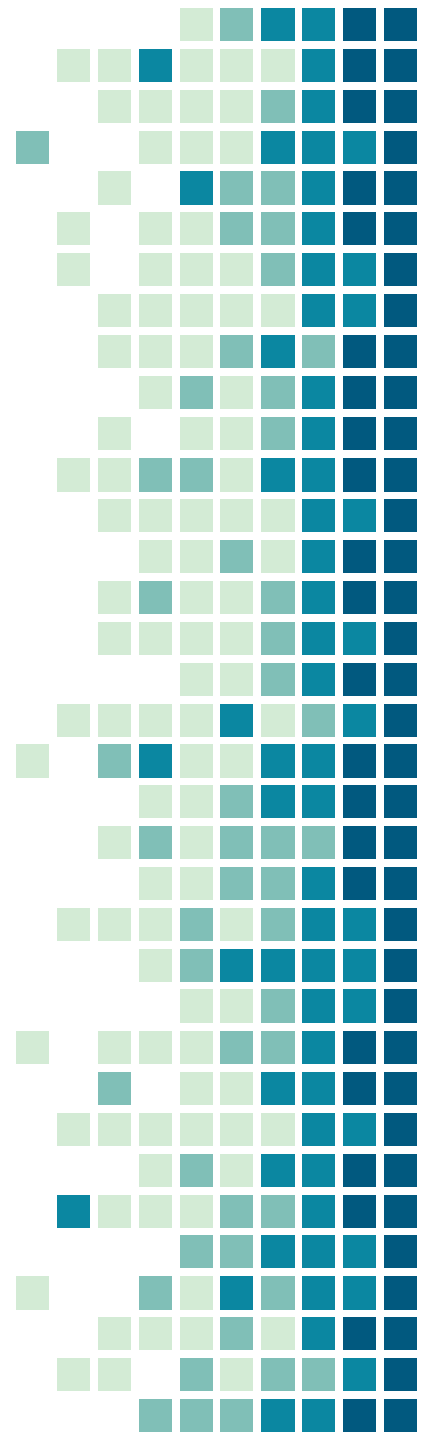
Old way of working



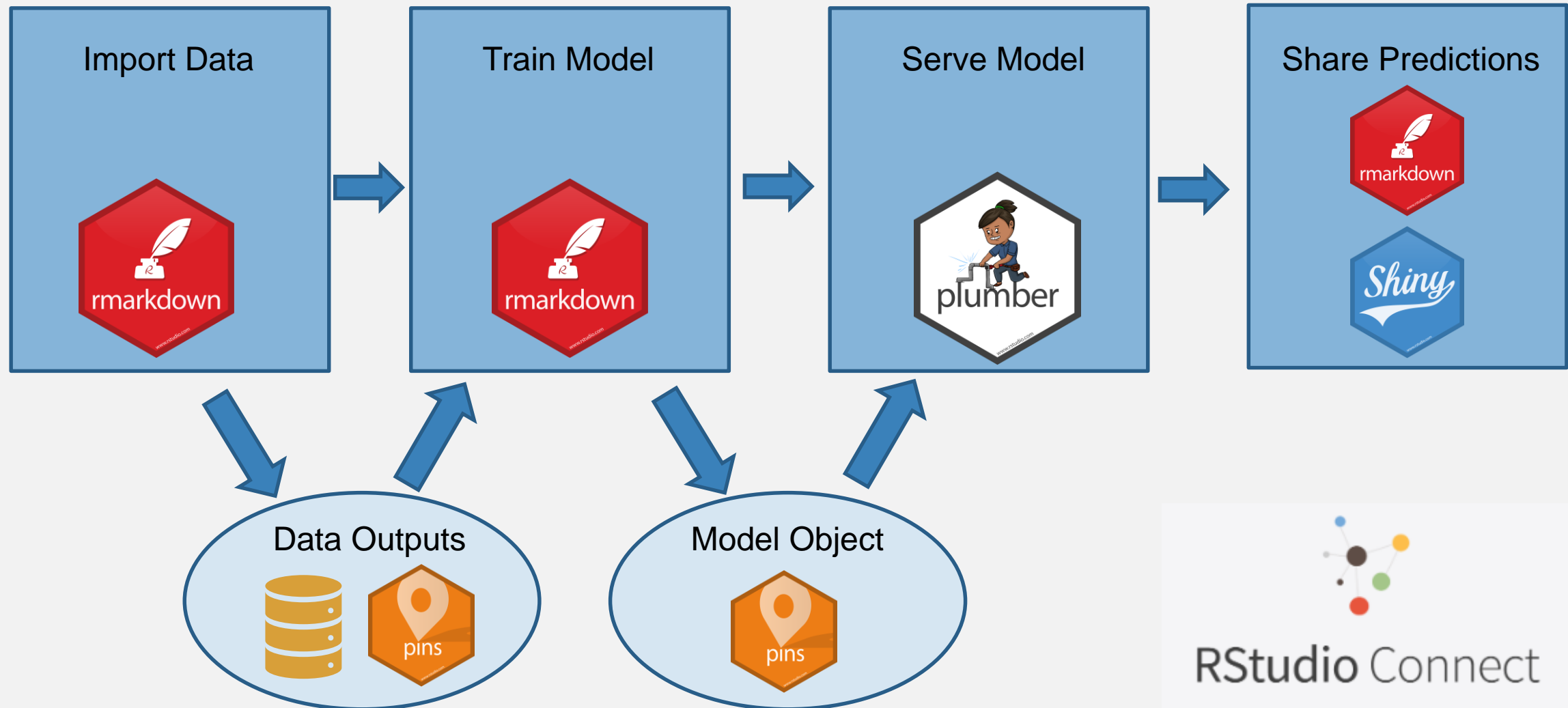
- Runs locally or scheduled on a server

Why something new?

- Easy to collaborate with team members
- Easy to schedule and maintain
- Accessible to business partners
- More reproducible
- Better Security



New way of working





R Markdown Overview

- R Markdown documents are flexible frameworks for data science
 - Execute and Save Code
 - Generate static and dynamic reports
- .Rmd files to execute code chunks, display text, and include YAML metadata to help the R Markdown build process
- Supported for SQL, R, and Python
- Reproducible, easy to track in Git, and easy to deploy

“ - *R Markdown Example*





R Markdown Best Practices

- Tips to make your R Markdowns functional and reproducible
 - Use parameters or dynamic variables rather than hardcoded variables
 - Name your code chunks
 - Modularize your code chunks
 - Don't use plain text passwords
 - Don't include unused packages, code, or dataframes
 - Use relative file paths
 - Use lintr and styler packages to point out style violations and automatically reformat code
 - Use testthat to incorporate unit tests
 - Convert your project to a package 😊



Pins Overview

- Use the pins package to pin remote resources locally, work offline, or cache results
- Share resources in local folders, RStudio Connect, Github, etc.
- Take an R object, dataframe, model object, etc. and pin to a board
- Great for collaboration
 - Instead of emailing objects or saving it to a database → share the name of the board and the pin name
 - Users will always get the newest version

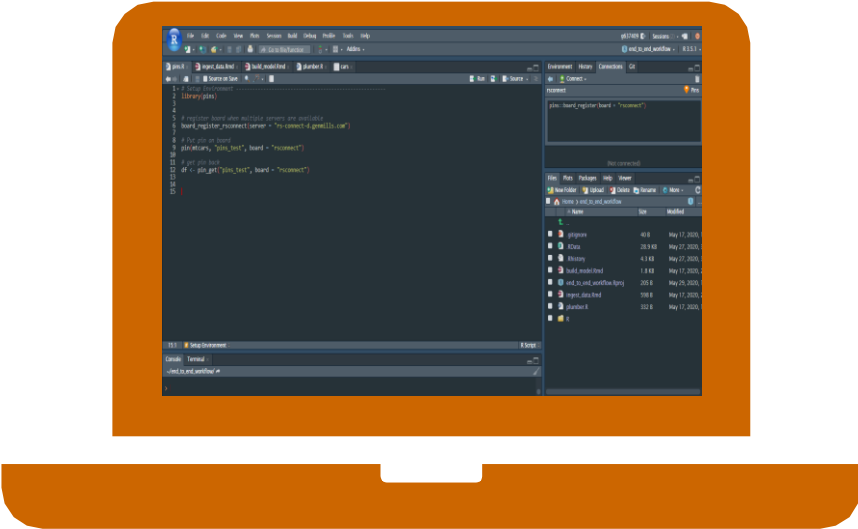


> pins::pin

Board

- RStudio Connect
- Github
- Azure
- GCP
- Many others!

> pins::pin_get



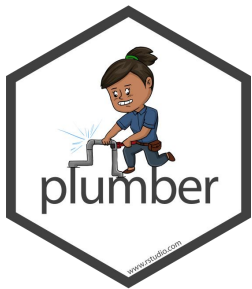
“ Pins Example





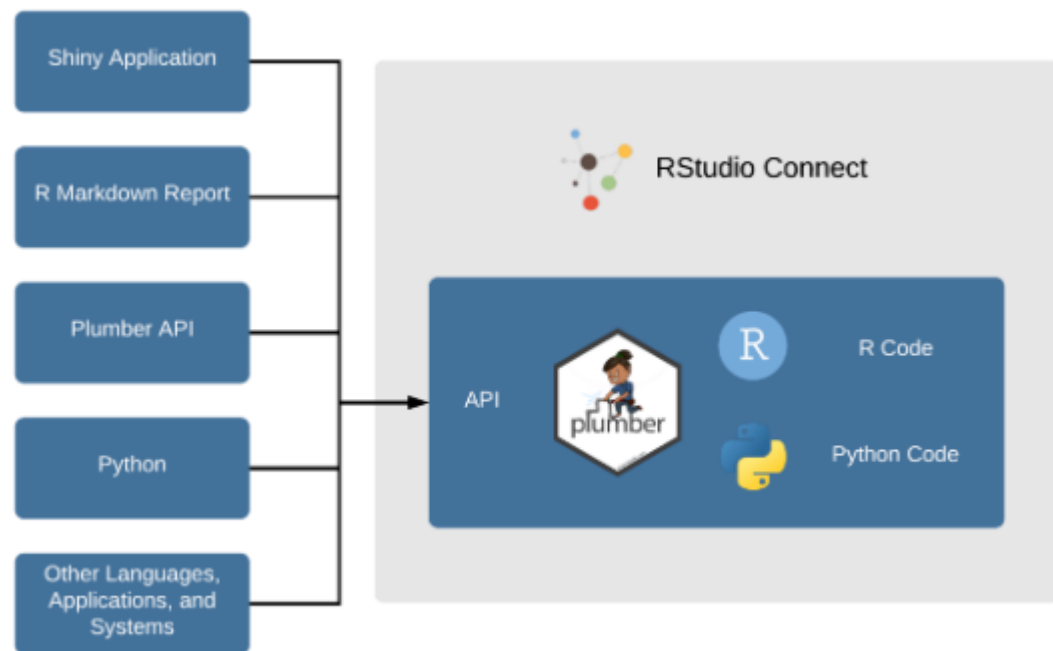
Pins Best Practices

- Pins is not a database replacement
- Pins are:
 - Small – 1GB or Less
 - Re-used
 - Re-created
 - Current – you just want the newest version
 - Shareable (+ you can secure them)
 - Great for objects that are computationally expensive and you'd like to cache the results
 - Great for model objects – simple to update the model independently from an API that serves model predictions



Plumber Overview

- Deploy machine learning models as a RESTful API built in R
- You can log details about API requests and responses
- You can call Plumber APIs from other languages



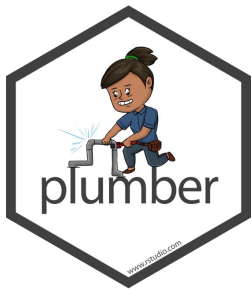
```
library(plumber)
```

```
## Return the sum of two numbers
## @param a The first number to add
## @param b The second number to add
## @post /sum
function(a, b) {
  as.numeric(a) + as.numeric(b)
}
```

```
plumb(file = "overview/plumber.R")$run()
```

“ ▪ *Plumber Example*

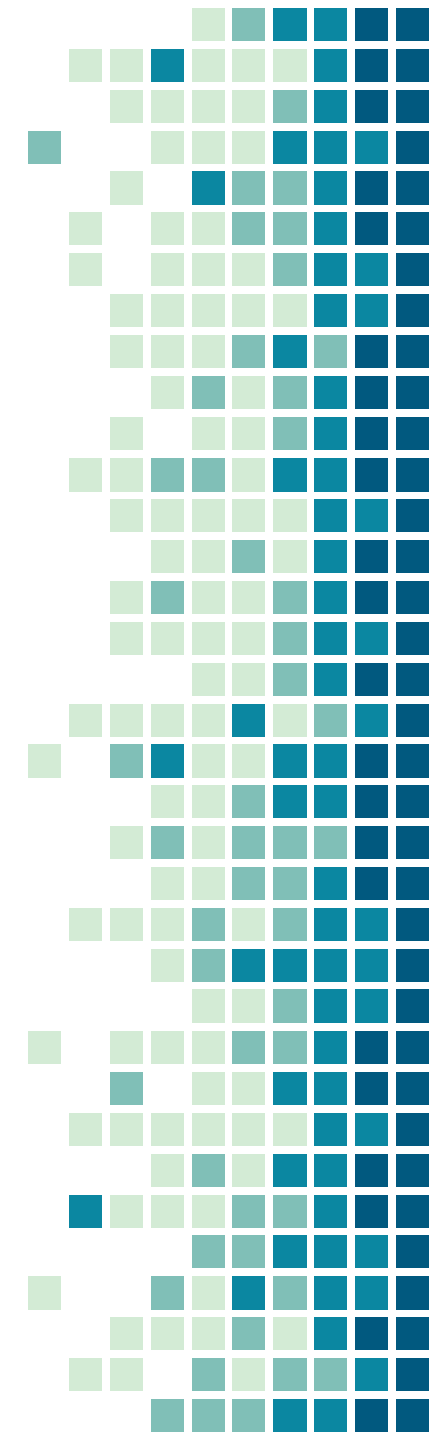




Plumber + Shiny??



- Plumber is great to host model APIs and get predictions
 - Shiny is intended for human consumption
 - Shiny will scale well as the data is not duplicated in memory for each user
 - The data is housed in one location (Plumber API & pins)
- The underlying data can be large, which is not suitable for a shiny app
 - The API can be updated without redeploying the shiny application.





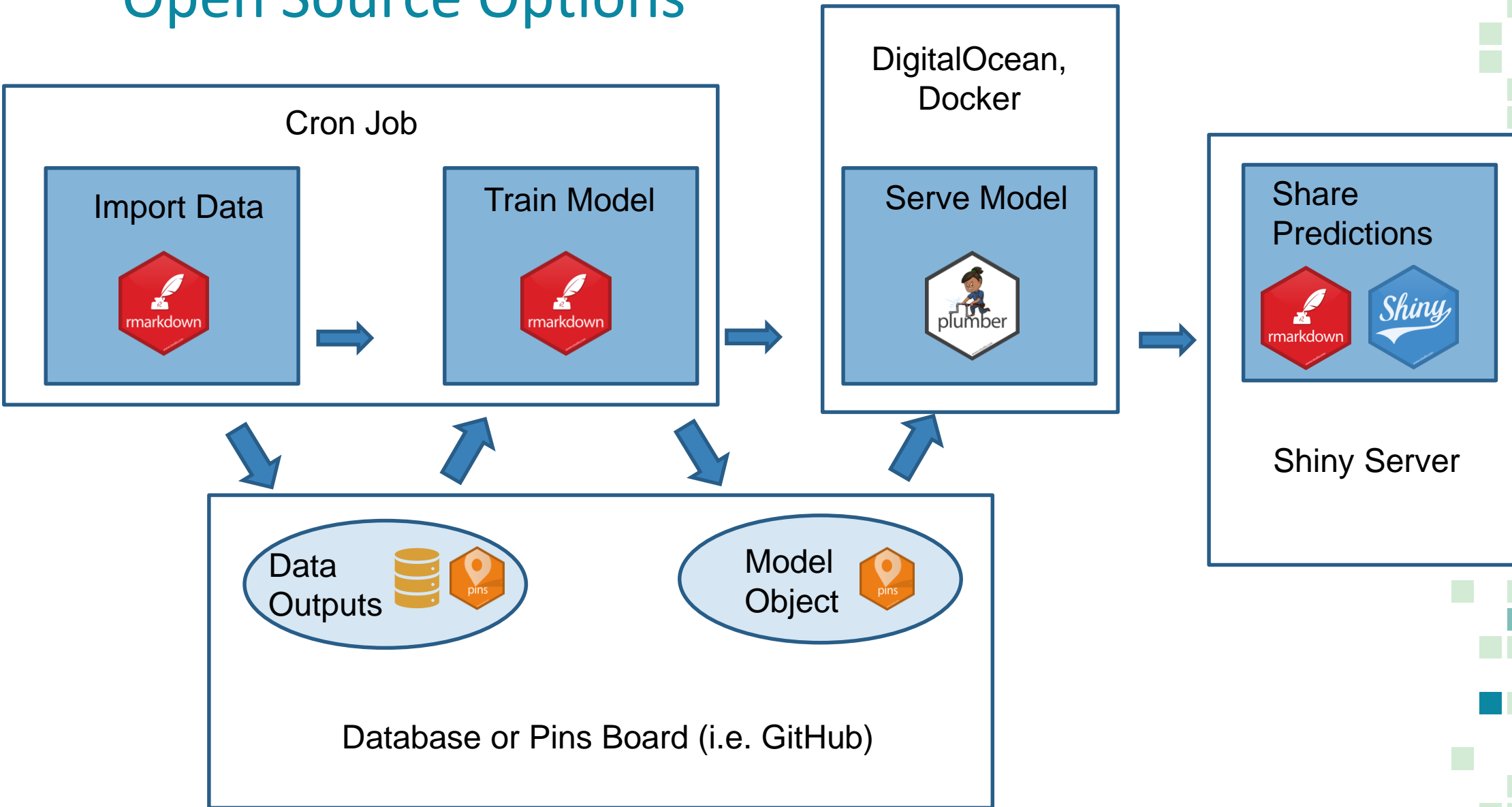
Shiny Overview

- A web application framework for R that allows you to turn your data into an interactive web app
 - User interface script that controls the layout and appearance by converting R code into HTML
 - Server script that contains the code needed to build the app
- Connect Shiny to Plumber APIs to get predictions from your model or monitor model performance

“ ▪ *Shiny Example*



Open Source Options





- Finally, remember to use **testthat** in your workflows!

WE'RE PASSIONATE ABOUT FOOD AND DRIVEN BY DATA

We need talent in technology to help us reshape the future.

Do you like to solve problems and make an impact? Join us and you'll work with talented, inclusive teams and disruptive technology to help impact the future.



BRING YOUR PASSION,

innovative ideas and expertise for
data, analytics and technology to
help us reshape the future.

Technology and data are deeply embedded in our business partnerships. They help bring the right data and solutions to the table and play an important role in better serving the world by making food people love. Our leaders are engaged and deeply committed to our success and our people are at the heart of everything we do. We take learning and personal growth seriously and recognize the importance of work-life agility.

Bring your expertise and problem-solving skills to our team and make an impact.

- We are hiring! Visit <https://careers.generalmills.com/datatechjobs/> to learn more!

Helpful Links

- R Markdown: <https://rmarkdown.rstudio.com/>
- Pins: <http://pins.rstudio.com/>
- Plumber: <https://www.rplumber.io/>
- Plumber: <https://rstudio.com/resources/webinars/plumbing-apis-with-plumber/>
- Rest APIs overview: <https://solutions.rstudio.com/examples/rest-apis-overview/>
- Shiny: <https://shiny.rstudio.com/>
- Testthat: <https://testthat.r-lib.org/>
- Putting R in Prod: <https://putrinprod.com/>
- Another End to End Example:
https://solutions.rstudio.com/tour/bike_predict/