R Shiny, Sparklyr, and RConnect

L.J. Arendse

Building ML pipelines, SparklyR usage and integration with Shiny, and deploying to RConnect

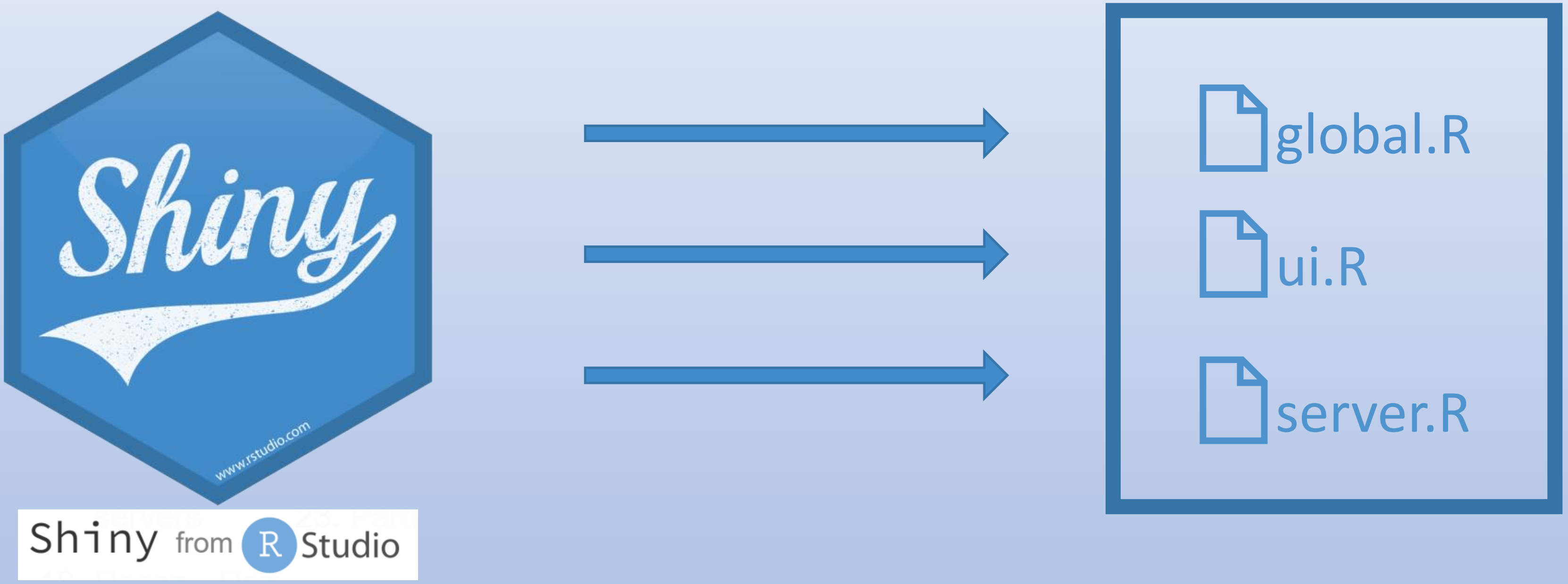
Shiny from R Studio

R Package: For rapid and easy development of interactive web apps

“Shiny combines the computational power of R with the interactivity of the modern web”

https://shiny.rstudio.com/

Shiny from R Studio

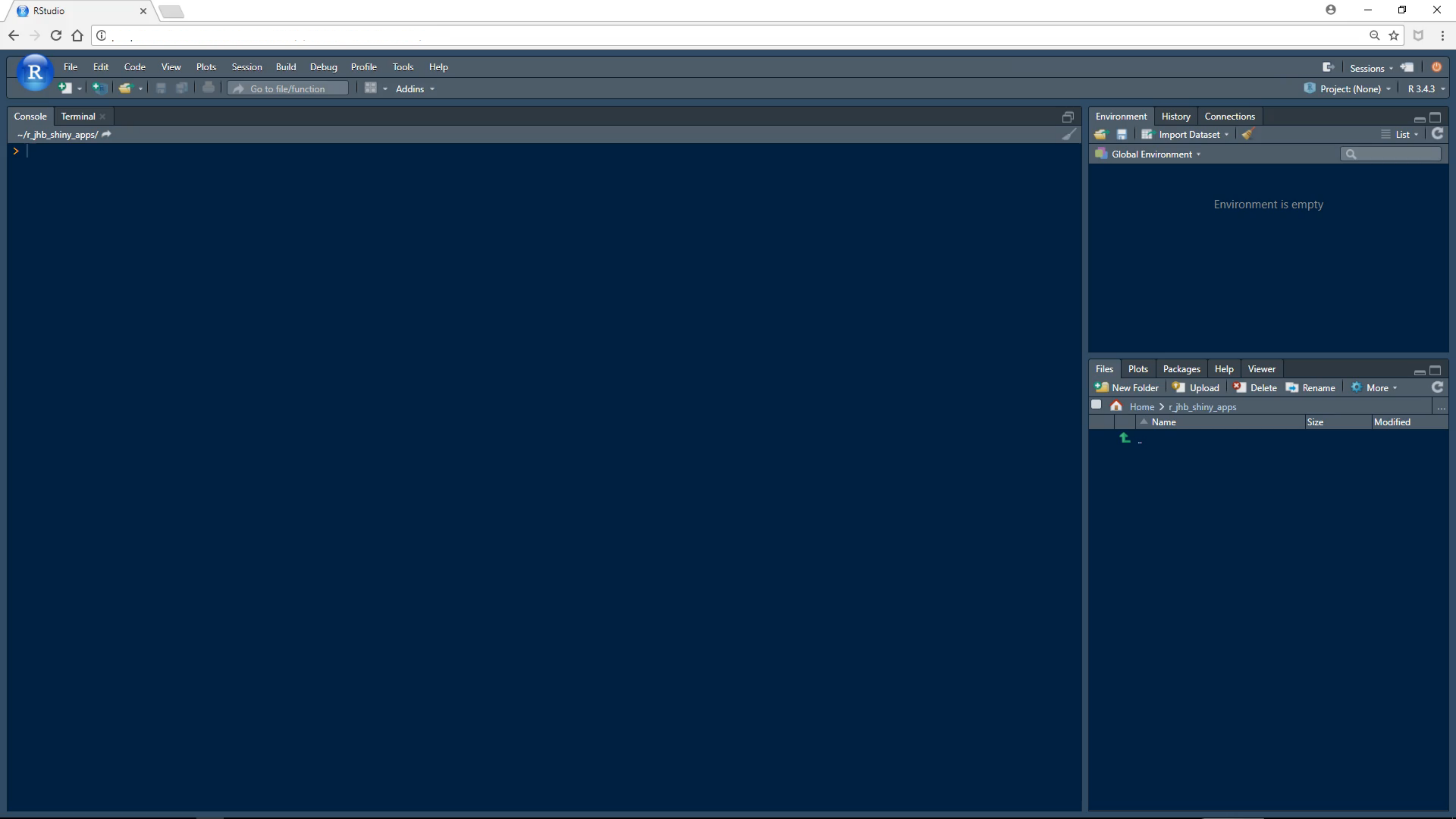


global.R

ui.R

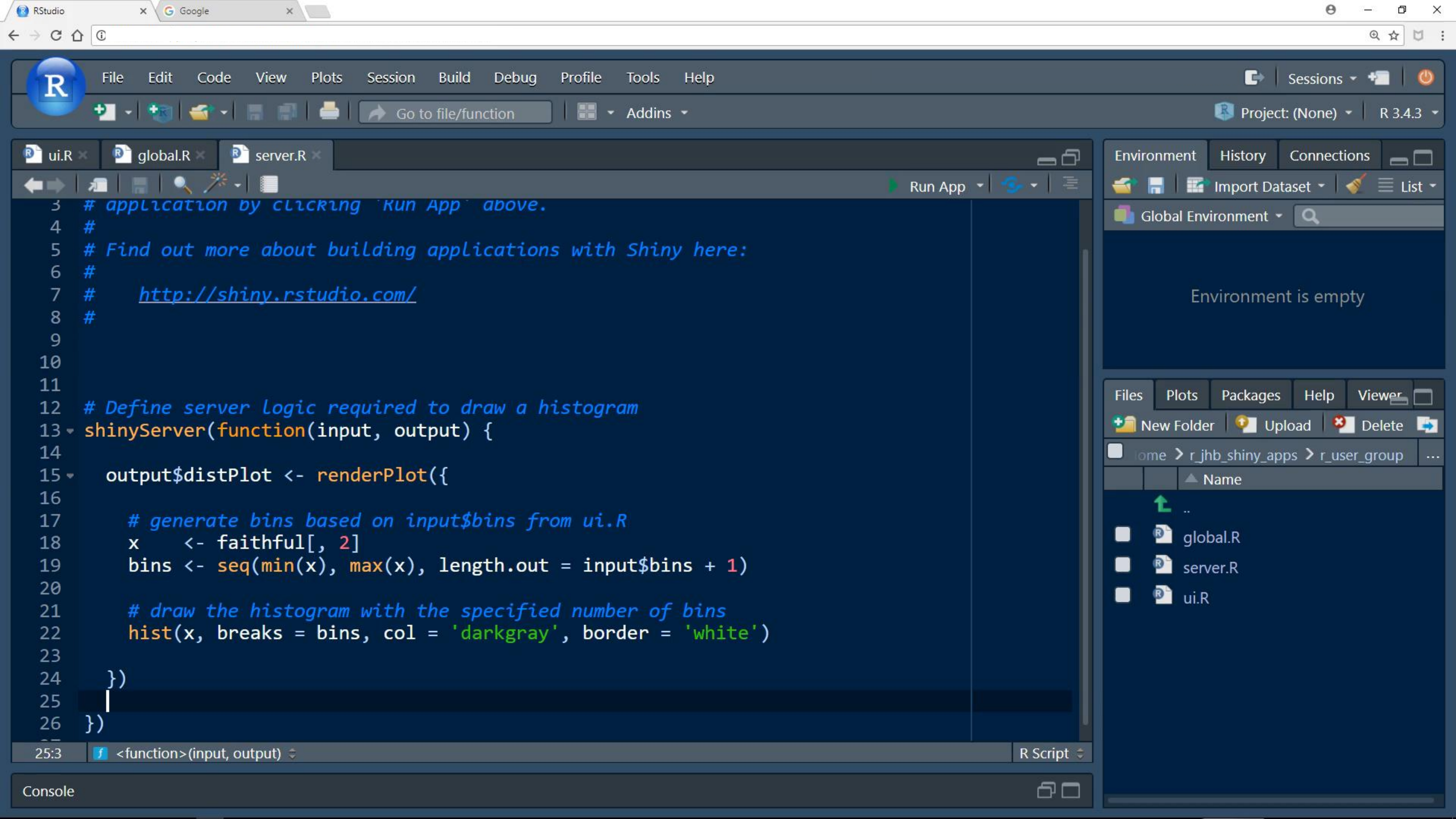
server.R

https://shiny.rstudio.com/



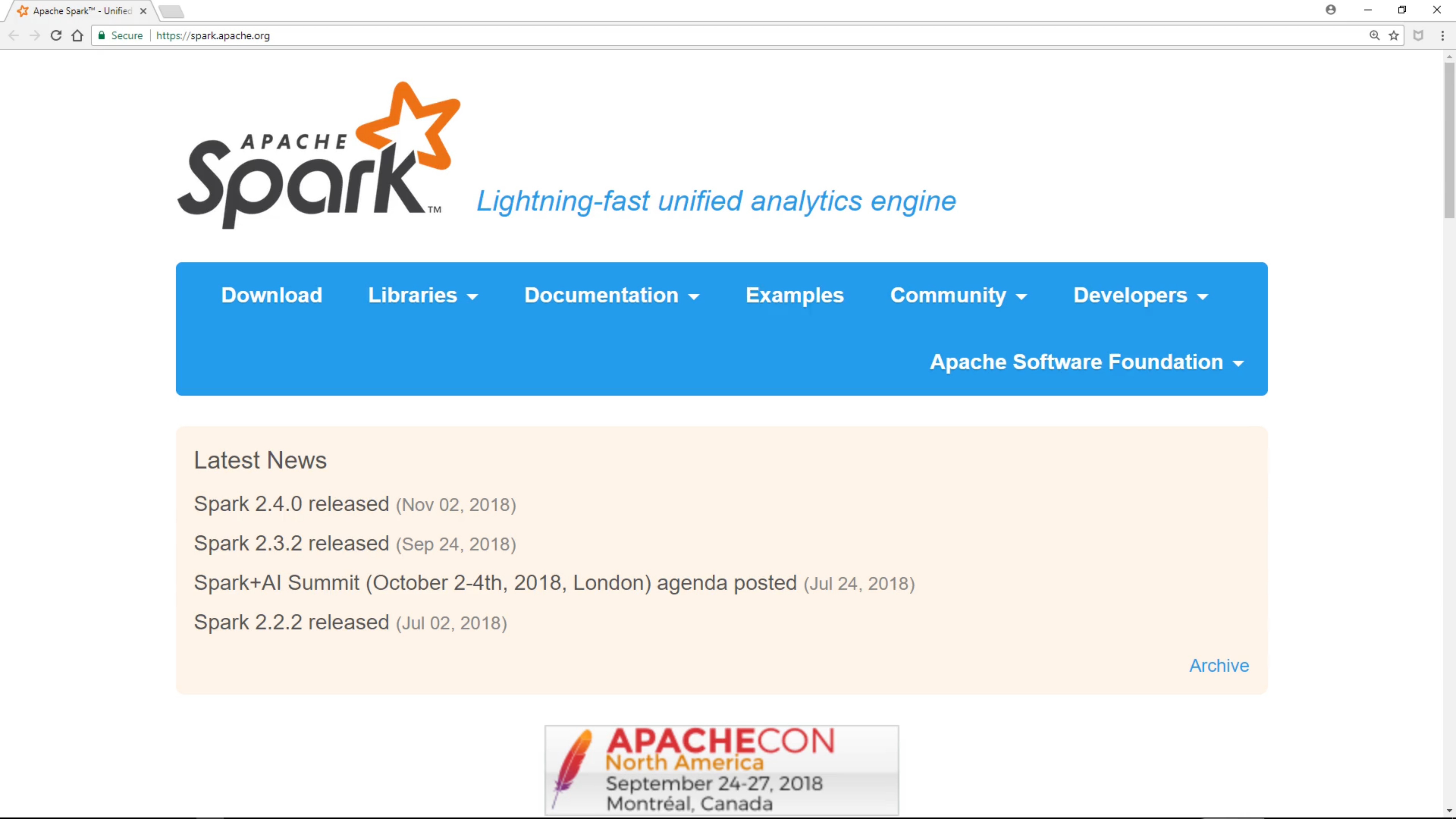
```
~/r_jhb_shiny_apps/
>
```

Environment is empty



```
# application by clicking 'Run App' above.
#
# Find out more about building applications with Shiny here:
#
#    http://shiny.rstudio.com/
#



# Define server logic required to draw a histogram
shinyServer(function(input, output) {

  output$distPlot <- renderPlot({

    # generate bins based on input$bins from ui.R
    x    <- faithful[, 2]
    bins <- seq(min(x), max(x), length.out = input$bins + 1)

    # draw the histogram with the specified number of bins
    hist(x, breaks = bins, col = 'darkgray', border = 'white')

  })

})
```

sparklyr
www.rstudio.com



Lightning-fast unified analytics engine

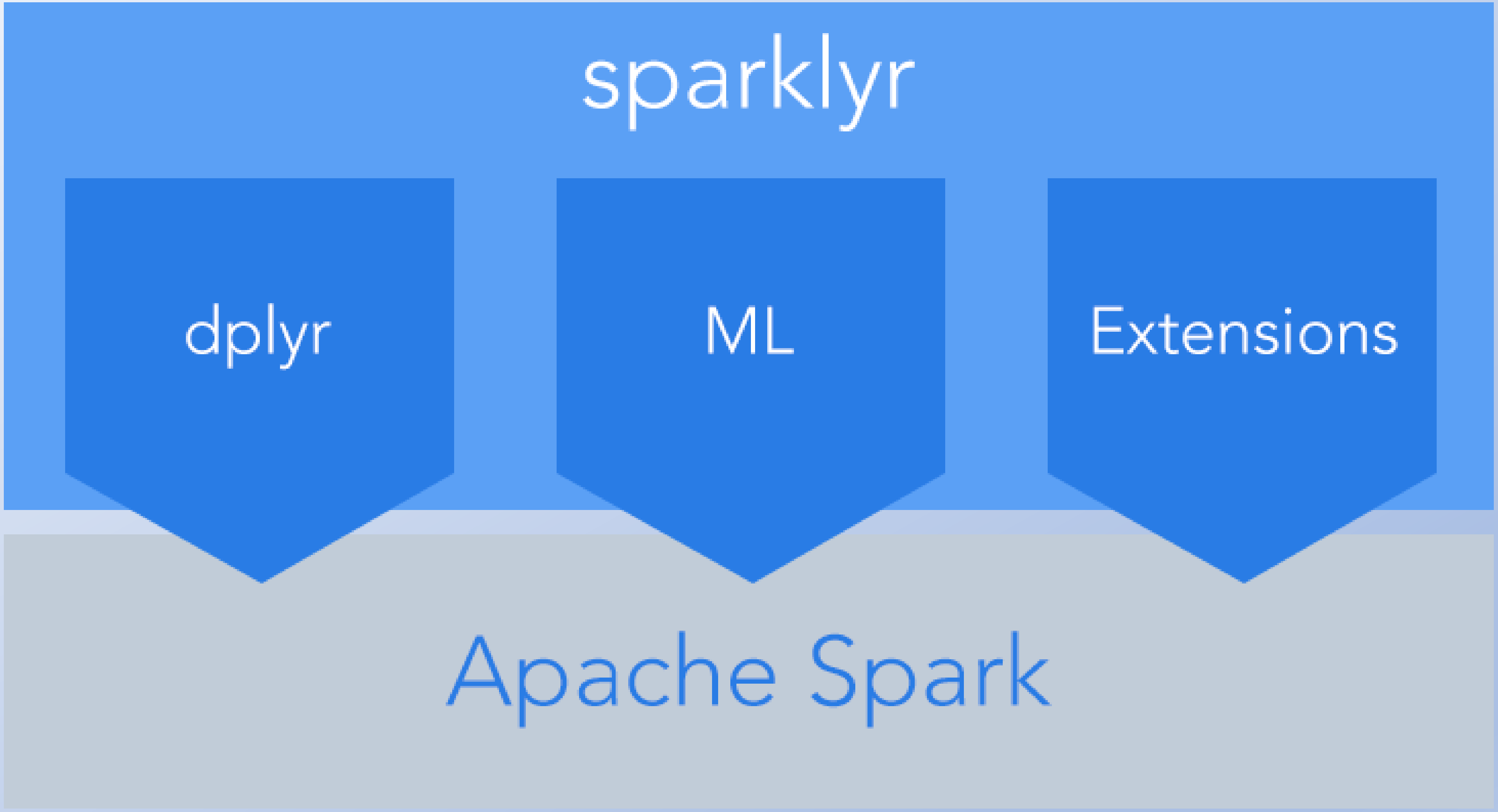
Latest News

Spark 2.4.0 released (Nov 02, 2018)

Spark 2.3.2 released (Sep 24, 2018)

Spark+AI Summit (October 2-4th, 2018, London) agenda posted (Jul 24, 2018)

Spark 2.2.2 released (Jul 02, 2018)



sparklyr

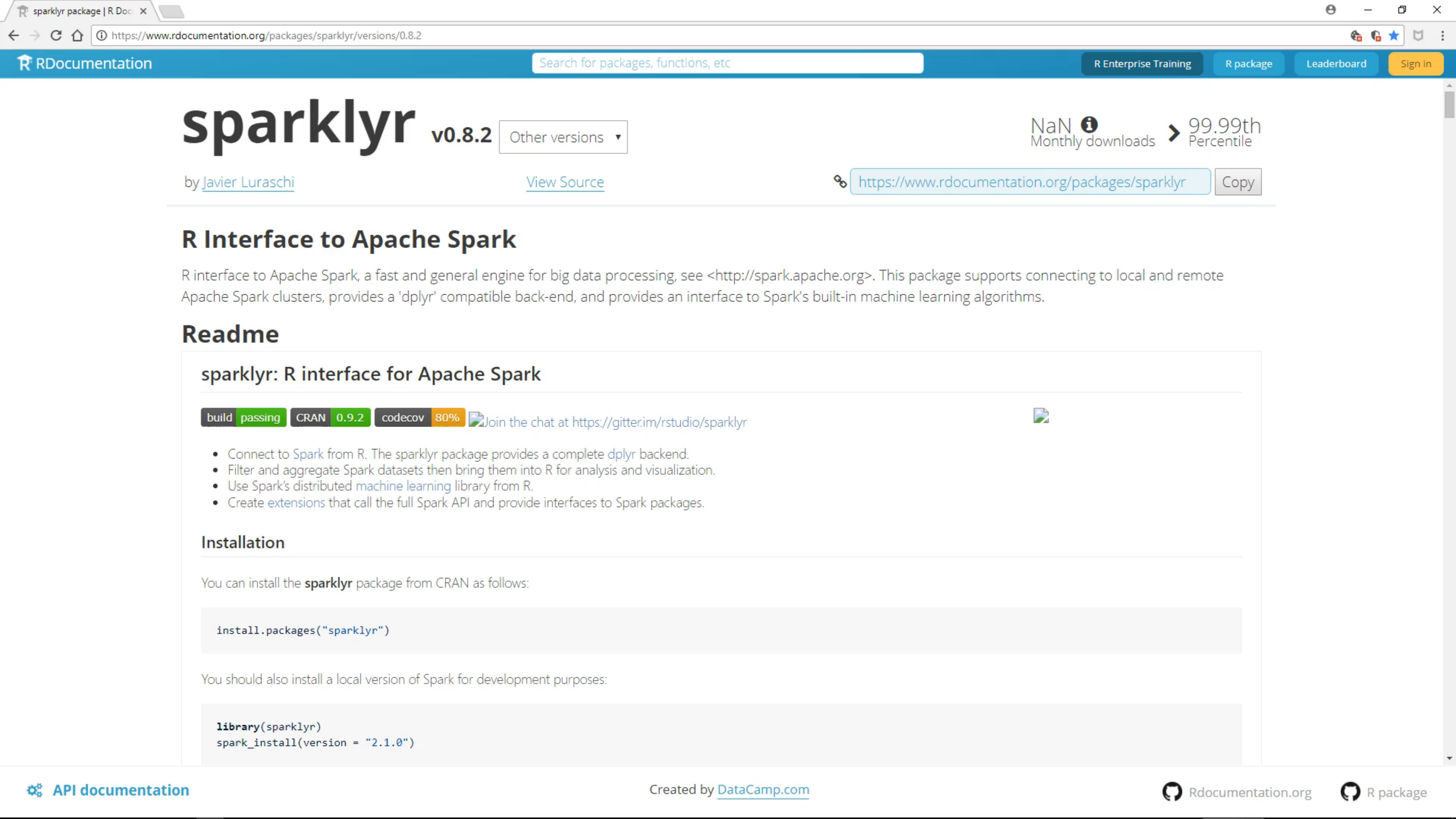
dplyr

ML

Extensions

Apache Spark

- https://www.rdocumentation.org/packages/sparklyr/versions/0.8.2
- https://spark.rstudio.com/



sparklyr v0.8.2

by Javier Luraschi

https://www.rdocumentation.org/packages/sparklyr

R Interface to Apache Spark

R interface to Apache Spark, a fast and general engine for big data processing, see <http://spark.apache.org>. This package supports connecting to local and remote Apache Spark clusters, provides a 'dplyr' compatible back-end, and provides an interface to Spark's built-in machine learning algorithms.

Readme

sparklyr: R interface for Apache Spark

Join the chat at https://gitter.im/rstudio/sparklyr

- Connect to Spark from R. The sparklyr package provides a complete dplyr backend.
- Filter and aggregate Spark datasets then bring them into R for analysis and visualization.
- Use Spark's distributed machine learning library from R.
- Create extensions that call the full Spark API and provide interfaces to Spark packages.

Installation

You can install the **sparklyr** package from CRAN as follows:

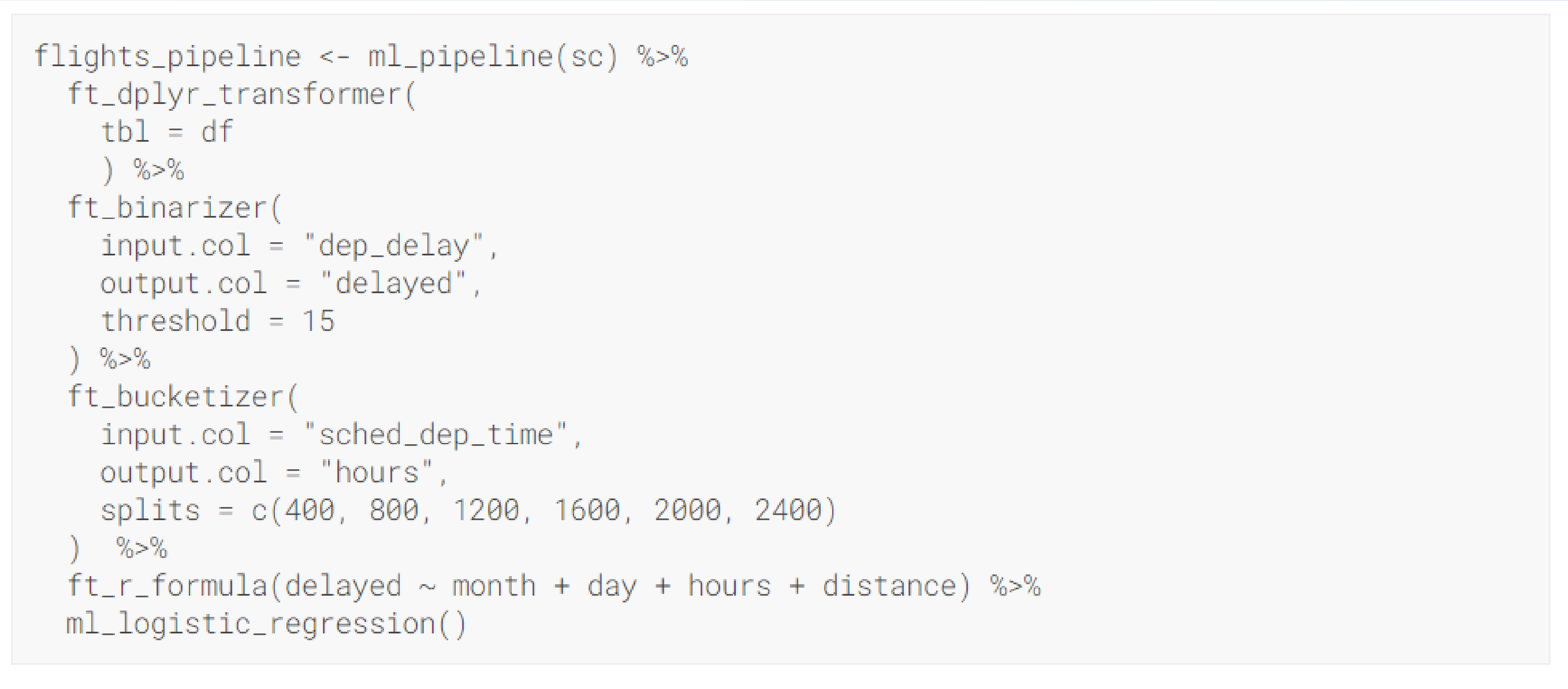
```
install.packages("sparklyr")
```

You should also install a local version of Spark for development purposes:

```
library(sparklyr)
spark_install(version = "2.1.0")
```

Created by DataCamp.com

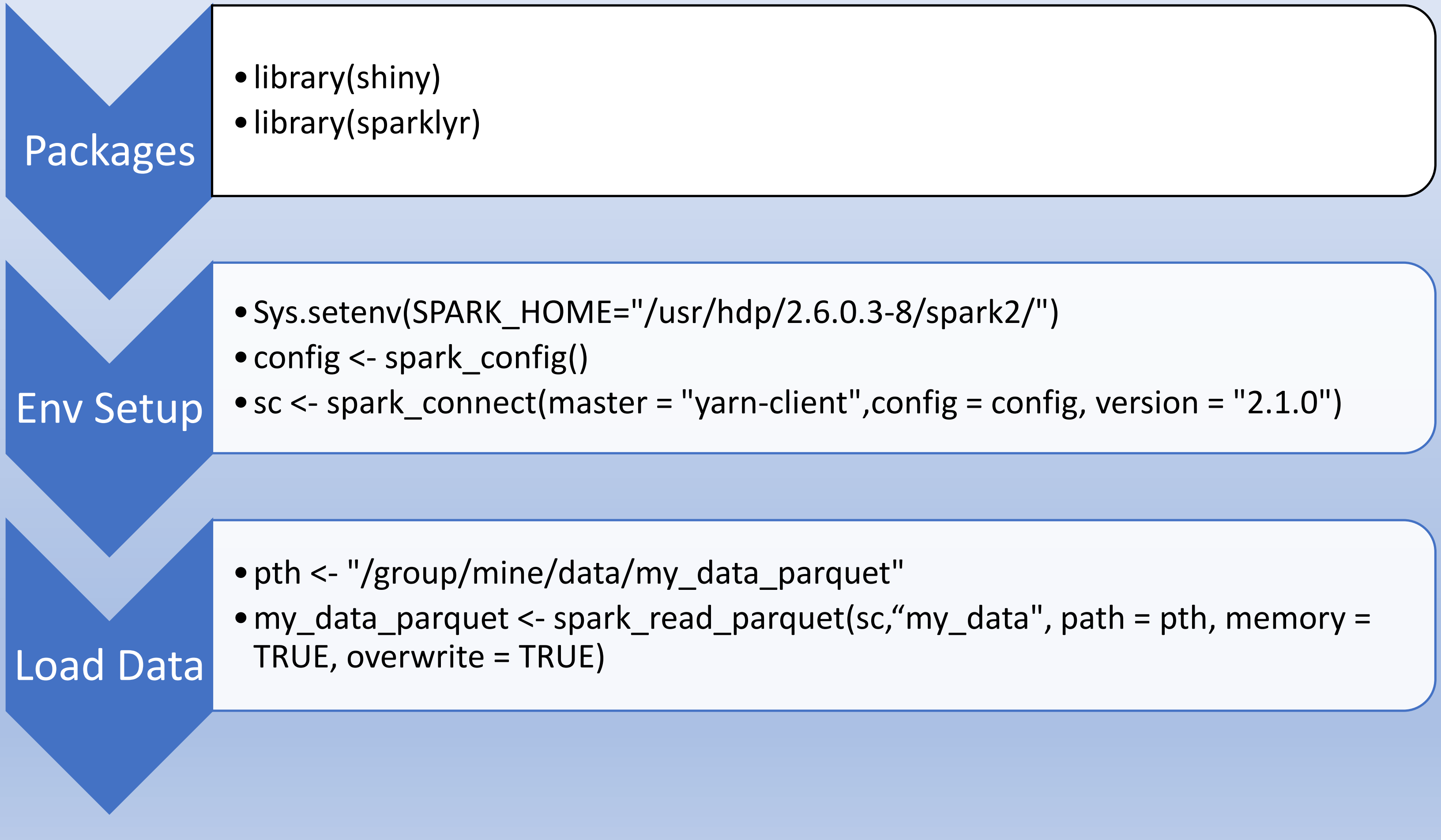
ML Pipelines



```r
flights_pipeline <- ml_pipeline(sc) %>%
  ft_dplyr_transformer(
    tbl = df
    ) %>%
  ft_binarizer(
    input.col = "dep_delay",
    output.col = "delayed",
    threshold = 15
  ) %>%
  ft_bucketizer(
    input.col = "sched_dep_time",
    output.col = "hours",
    splits = c(400, 800, 1200, 1600, 2000, 2400)
  )  %>%
  ft_r_formula(delayed ~ month + day + hours + distance) %>%
  ml_logistic_regression()
```

- https://spark.rstudio.com/guides/pipelines/

Shiny and SparklyR



Packages

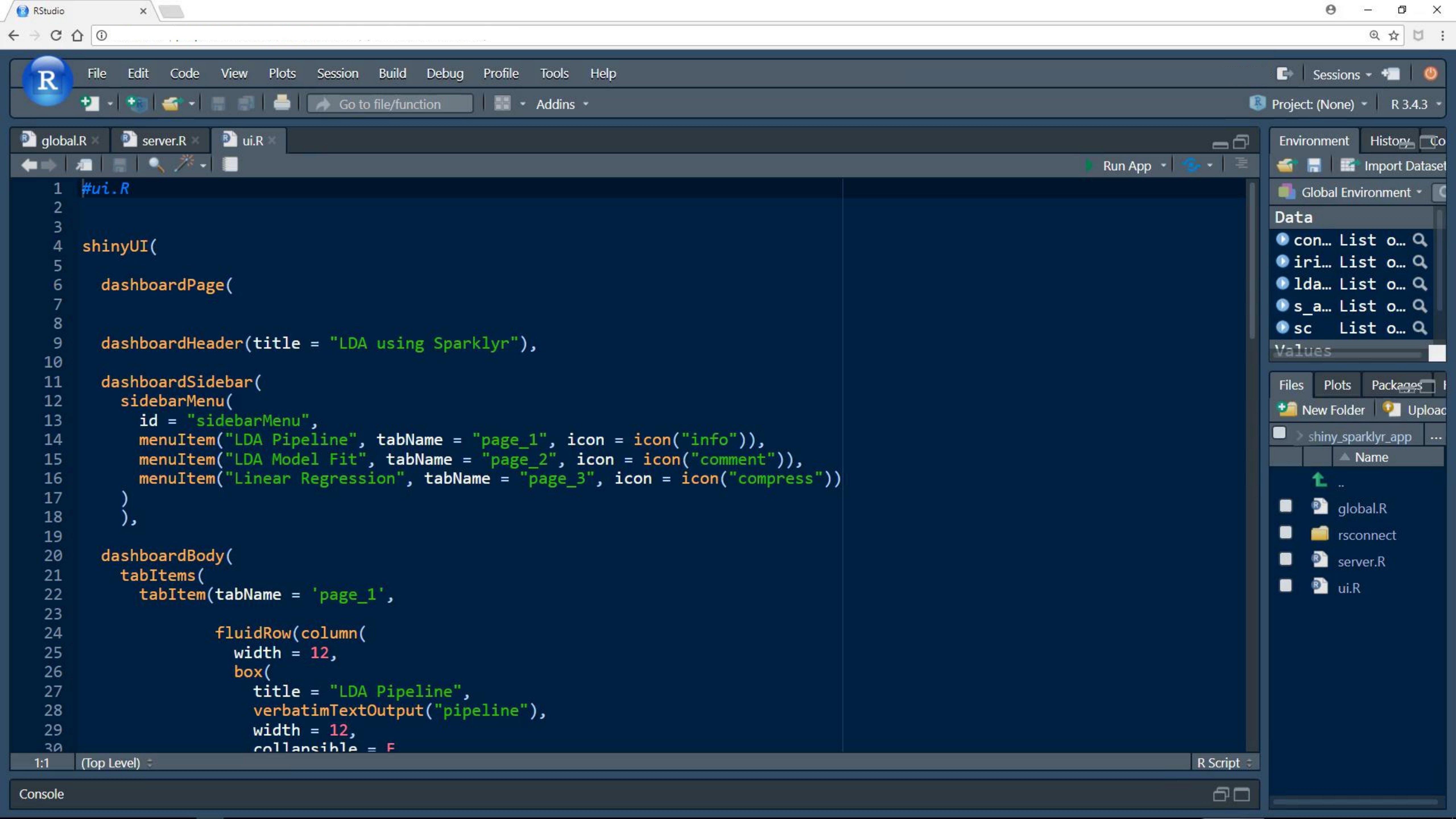
- library(shiny)
- library(sparklyr)

Env Setup

- Sys.setenv(SPARK_HOME="/usr/hdp/2.6.0.3-8/spark2/")
- config <- spark_config()
- sc <- spark_connect(master = "yarn-client",config = config, version = "2.1.0")

Load Data

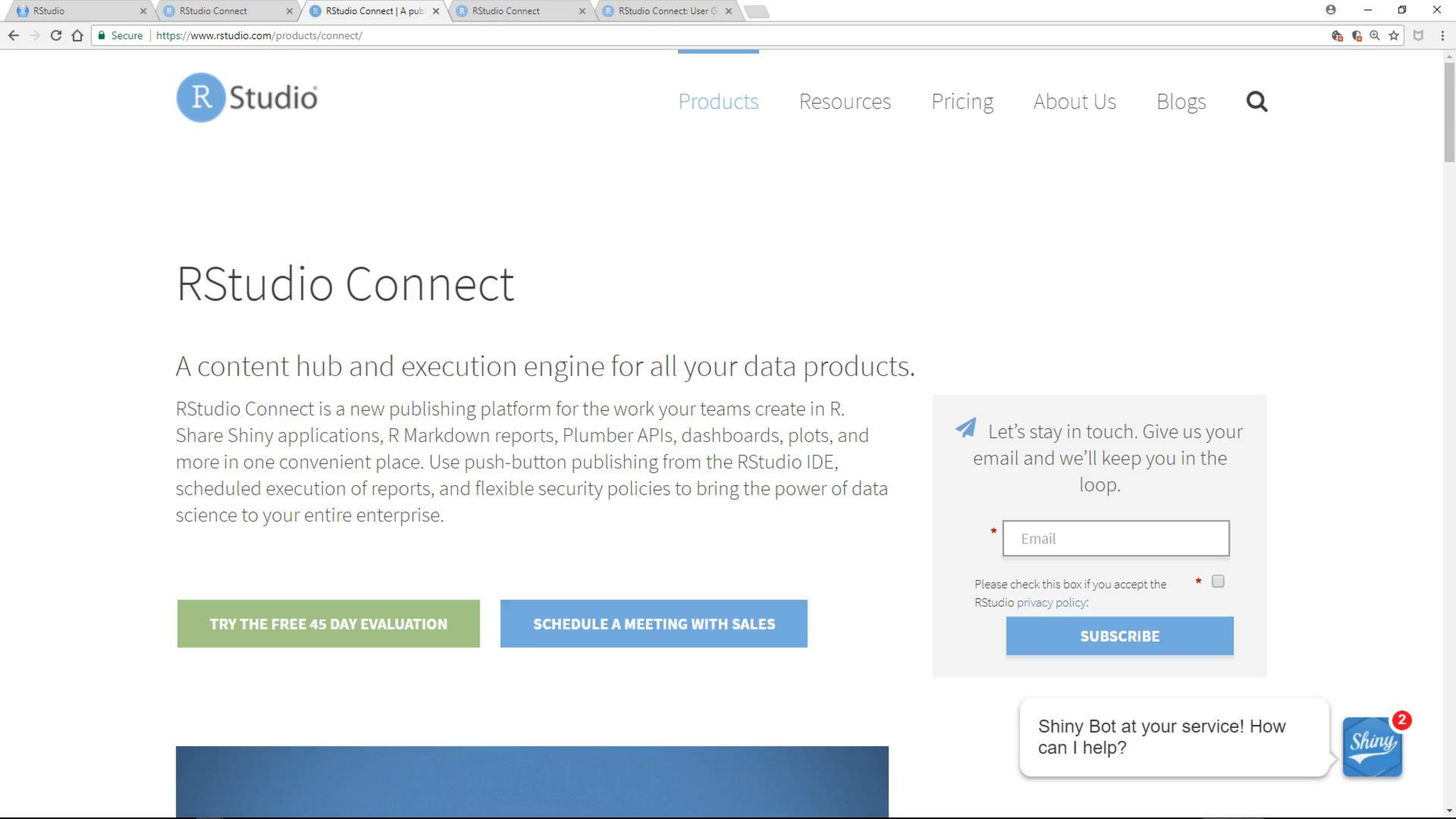
- pth <- "/group/mine/data/my_data_parquet"
- my_data_parquet <- spark_read_parquet(sc,“my_data", path = pth, memory = TRUE, overwrite = TRUE)



```
#ui.R


shinyUI(

  dashboardPage(


  dashboardHeader(title = "LDA using Sparklyr"),

  dashboardSidebar(
    sidebarMenu(
      id = "sidebarMenu",
      menuItem("LDA Pipeline", tabName = "page_1", icon = icon("info")),
      menuItem("LDA Model Fit", tabName = "page_2", icon = icon("comment")),
      menuItem("Linear Regression", tabName = "page_3", icon = icon("compress"))
    )
    ),

  dashboardBody(
    tabItems(
      tabItem(tabName = 'page_1',

               fluidRow(column(
                width = 12,
                box(
                 title = "LDA Pipeline",
                 verbatimTextOutput("pipeline"),
                 width = 12,
                 collapsible = F
```

RStudio Connect

“Distribute data science across your organisation”



RStudio Connect

A content hub and execution engine for all your data products.

RStudio Connect is a new publishing platform for the work your teams create in R. Share Shiny applications, R Markdown reports, Plumber APIs, dashboards, plots, and more in one convenient place. Use push-button publishing from the RStudio IDE, scheduled execution of reports, and flexible security policies to bring the power of data science to your entire enterprise.

TRY THE FREE 45 DAY EVALUATION

SCHEDULE A MEETING WITH SALES

Let's stay in touch. Give us your email and we'll keep you in the loop.

* Email

Please check this box if you accept the RStudio privacy policy: * □

SUBSCRIBE

Shiny Bot at your service! How can I help?

Thank you

Questions?