



# HANDS-ON LAB SESSION SJSU

## Data Science Experience Desktop IBM

**Abstract**  
Use cases for using RStudio on DSX Desktop

## Table of Contents

<b>Requirements:</b> .....	<b>3</b>
<b>Exercise 1: Create RStudio session</b> .....	<b>3</b>
<b>Exercise 2: Loading packages in RStudio</b> .....	<b>4</b>
<b>Exercise 3: Adding data files in RStudio</b> .....	<b>5</b>
<b>Exercise 4: Setup Git on RStudio and Associate with GitHub</b> .....	<b>7</b>
<b>Exercise 5: Using Shiny apps</b> .....	<b>10</b>
<b>Exercise 6: Access IBM Analytics for Apache Spark from RStudio</b> .....	<b>15</b>

## Requirements:

This lab requires the user to install IBM Data Science Experience Desktop (IBM DSX Desktop) on their laptop. The latest version can be downloaded from <https://datascience.ibm.com/desktop>

Get familiar with IBM Data Science Experience RStudio by learning about the product features and functionality in the hand-on lab.

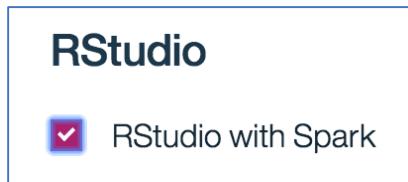
### Exercise 1: Create RStudio session

Tasks:

- Choose RStudio with spark image while installing IBM DSX Desktop
- Load RStudio

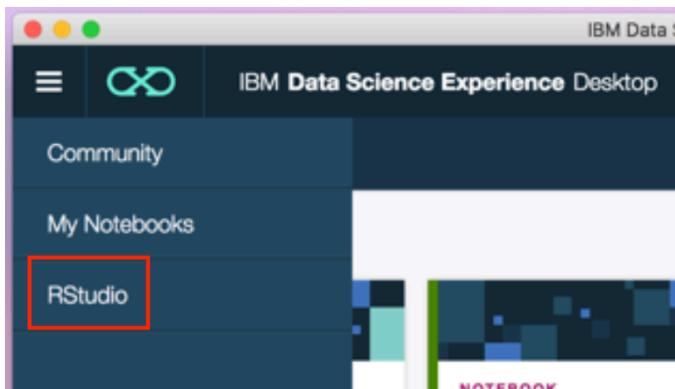
Description:

- Follow instructions in <https://datascience.ibm.com/docs/content/desktop/install.html> to set up RStudio
- Choose RStudio with spark while on the Settings screen during installation



- Where to find RStudio?

In the DSX Desktop side menu, click on the RStudio link.



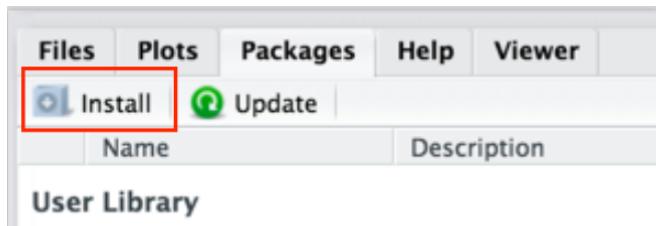
## Exercise 2: Loading packages in RStudio

### Tasks:

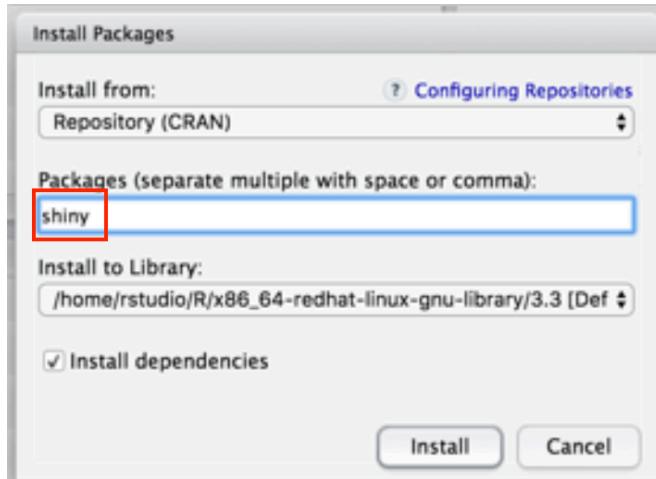
- View and load default system R packages
- Install and load user R packages

### Description:

- Select *View -> Show Packages* menu. This will open view Packages and then scroll to **System Library**. These are the R packages by default available for this R session. Select the packages to load to active R session.
- Also you can install an R package by using the RStudio interface. To do this, click **Packages** in the bottom right section of RStudio, then click **Install** below it. These installed R packages located in **User Library** section.



- Next, type the names of the packages that you want to install in the pop-up, and click **Install**.



- If you want to install packages from within your script, use the following code where "myPackage" is replaced with the name of your package.

```
install.package("myPackage")
library(myPackage)
```

All packages installed by user locate at "/home/R/x86\_64-redhat-linux-gnu-library/3.3" folder.

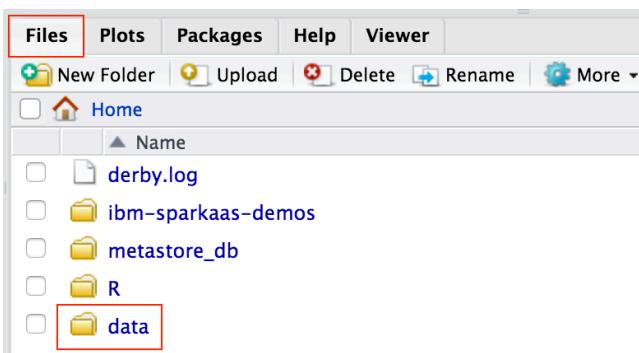
## Exercise 3: Adding data files in RStudio

### Tasks:

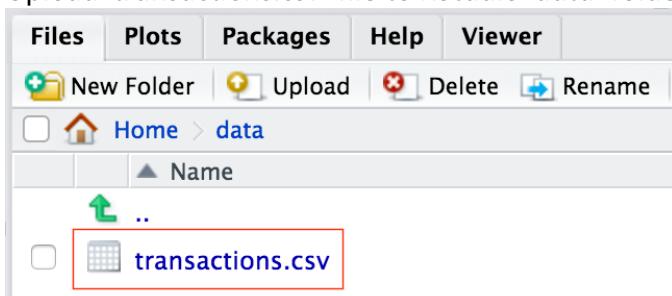
- View file browser, create data directory
- Upload 'transactions.csv' file to data directory
- View and use transactions.csv file

### Description:

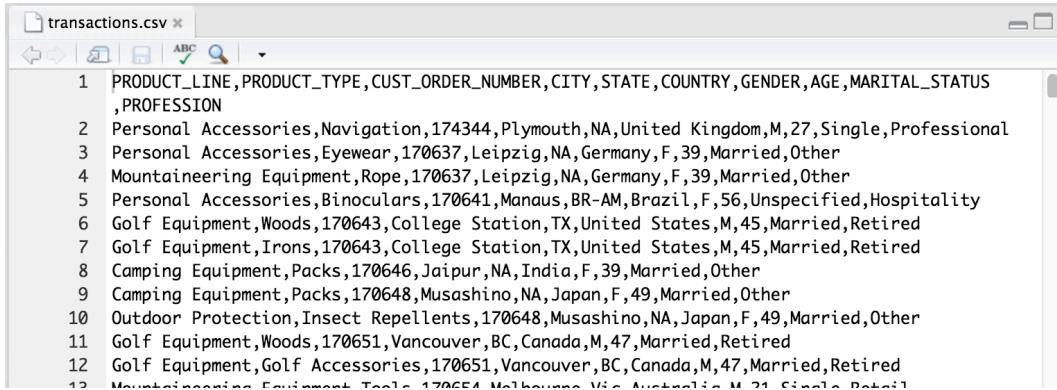
1. Select 'View->Show Files' menu. This will open file browser. Here you can navigate RStudio file system, upload files from local file system and other file operations.
2. Select 'New Folder' toolbar to create folder 'data'



3. Get 'transactions.csv' file from 'IBMDatascience' below link and store to local file system  
<https://raw.githubusercontent.com/IBMDatascience/datafirst/master/datascientist/machinelearning/data/transactions.csv>
4. Upload 'transactions.csv' file to RStudio 'data' folder



5. Because the files that you added are now in the server file structure, you will read them from there. After the file is uploaded, click the file name to see a preview in a data viewer.



1	PRODUCT_LINE,PRODUCT_TYPE,CUST_ORDER_NUMBER,CITY,STATE,COUNTRY,GENDER,AGE,MARITAL_STATUS,PROFESSION
2	Personal Accessories,Navigation,174344,Plymouth,NA,United Kingdom,M,27,Single,Professional
3	Personal Accessories,Eyewear,170637,Leipzig,NA,Germany,F,39,Married,Other
4	Mountaineering Equipment,Rope,170637,Leipzig,NA,Germany,F,39,Married,Other
5	Personal Accessories,Binoculars,170641,Manaus,BR-AM,Brazil,F,56,Unspecified,Hospitality
6	Golf Equipment,Woods,170643,College Station,TX,United States,M,45,Married,Retired
7	Golf Equipment,Irons,170643,College Station,TX,United States,M,45,Married,Retired
8	Camping Equipment,Packs,170646,Jaipur,NA,India,F,39,Married,Other
9	Camping Equipment,Packs,170648,Musashino,NA,Japan,F,49,Married,Other
10	Outdoor Protection,Insect Repellents,170648,Musashino,NA,Japan,F,49,Married,Other
11	Golf Equipment,Woods,170651,Vancouver,BC,Canada,M,47,Married,Retired
12	Golf Equipment,Golf Accessories,170651,Vancouver,BC,Canada,M,47,Married,Retired
13	Mountaineering Equipment,Tools,170651,Melbourne,VA,Australia,M,21,Single,Retired

To access this file in R, set the working directory as the directory with the data set. You can do this by navigating to the directory with the file and clicking **More**, then click **Set as Working Directory**.

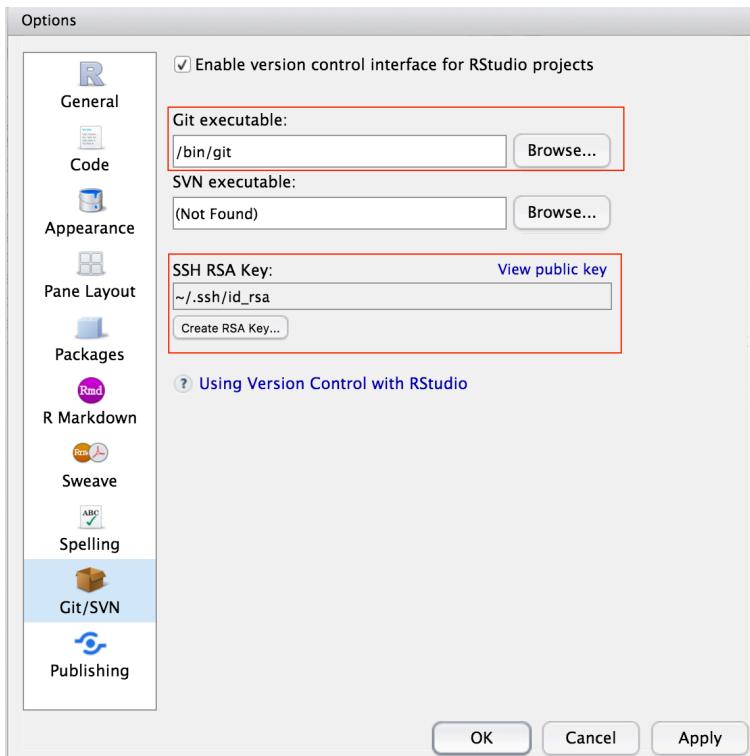
## Exercise 4: Setup Git on RStudio and Associate with GitHub

### Tasks:

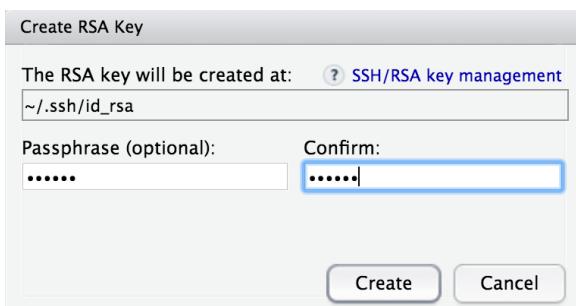
- Configure Git and create RSA key
- Create new project in RStudio and associate with Github

### Description:

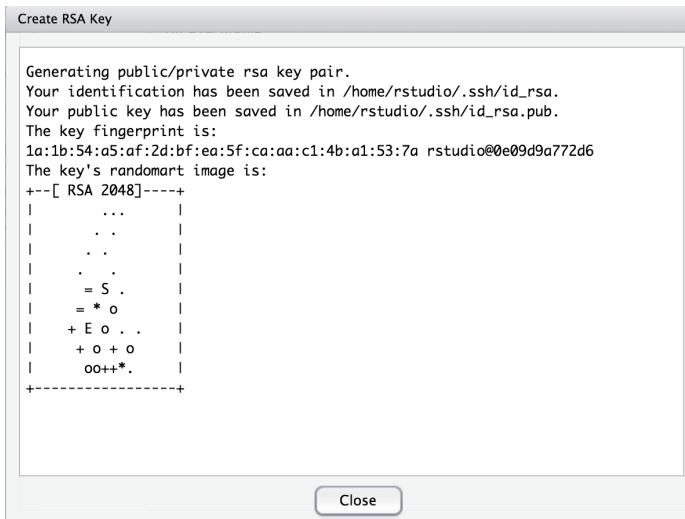
1. In RStudio, *Tools->Global Options*, select Git//SVN tab



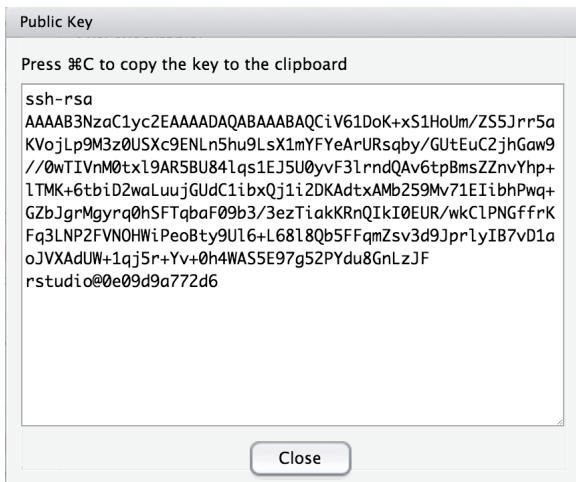
2. Now select, *Create RSA Key*, then enter passphrase like below:



3. This should generate RSA keys



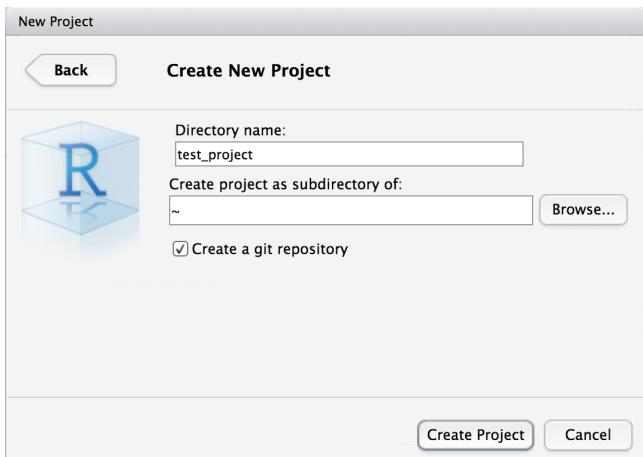
4. Now click *View public key*, and copy the displayed public key



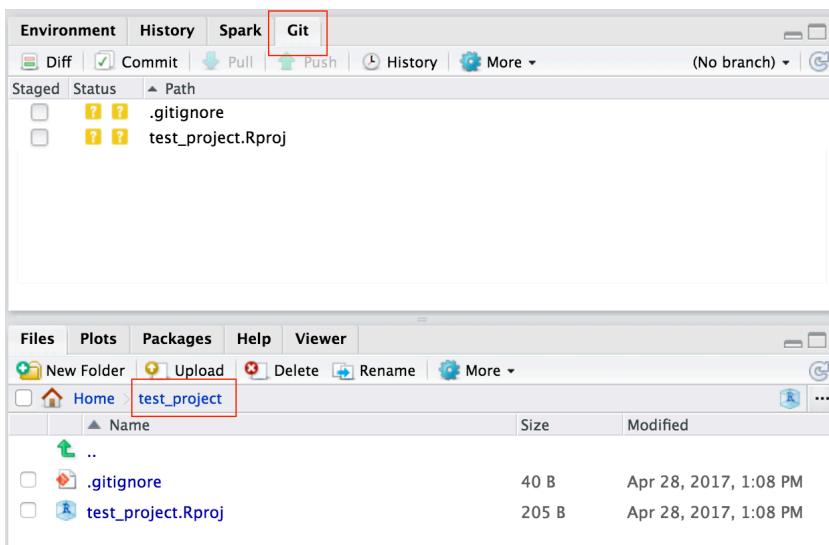
5. Open your Github account settings and click the SSH keys tab and Add SSH key. Paste in the public key you have copied from RStudio.
6. Configure git account email and username in RStudio console(View->Move Focus to Console)

```
system('git config --global user.email "user@email.com"')  
system('git config --global user.name "username"')
```

7. Create new project in RStudio  
*Create Project -> New Directory -> Empty Project*



8. This should create a test\_project in RStudio and create Github repo for this project in Github. See below:



## Exercise 5: Using Shiny apps

Tasks:

- Clone **dsx-shiny-apps** Github repo
- Save and upload flights.zip and predict\_accidents.zip shiny apps
- Run shiny app for Predictions of car accident in NYC based on weather data
- Run shiny app for Analyzing flight delays
- Run the Shiny App - Flex Dashboard

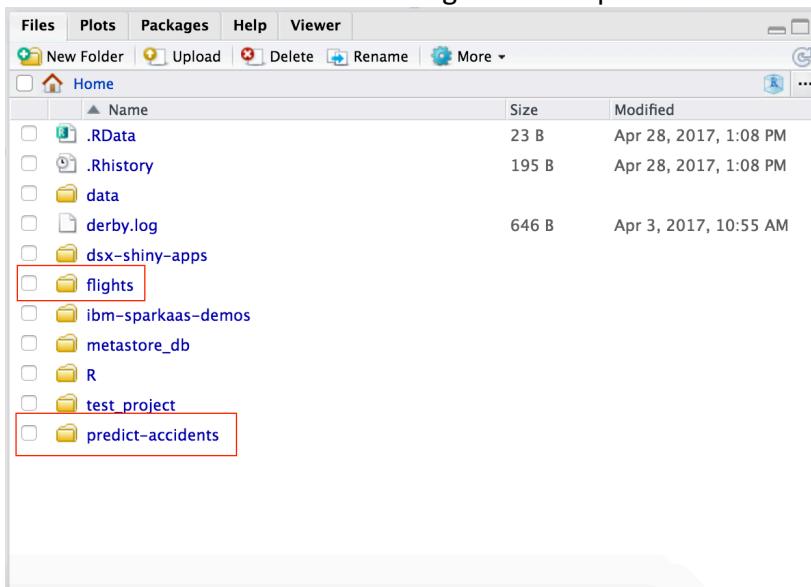
Description:

1. Download the zip files flights.zip and predict-accidents.zip from the below links:  
<https://github.com/IBMDatascience/dsx-shiny-apps/raw/master/flights.zip>  
<https://github.com/IBMDatascience/dsx-shiny-apps/raw/master/predict-accidents.zip>
2. Extract the zip files and copy the folders to the user volume on your laptop:  
For mac:  
Copy the folders to location:  
`/Users/<username>/Library/Application Support/ibm-dsx-desktop/RStudio/myRFiles/`

For Windows:

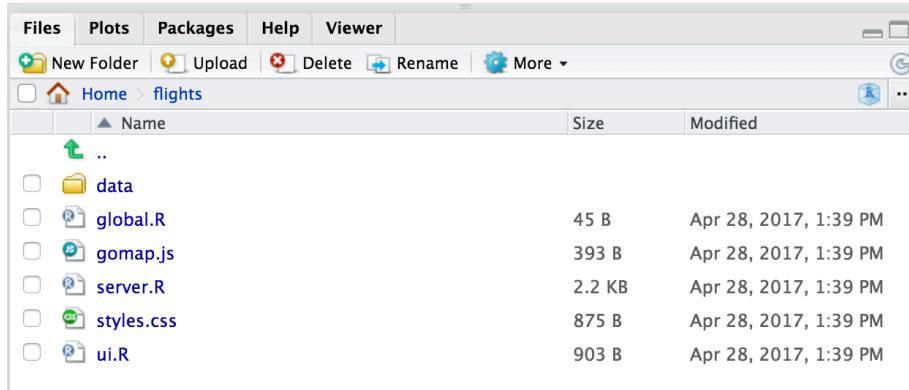
Copy the folders to location:  
`%APPDATA%\ibm-dsx-desktop\RStudio\myRFiles\`

3. You will see these folders in the right bottom pane in the Files section:

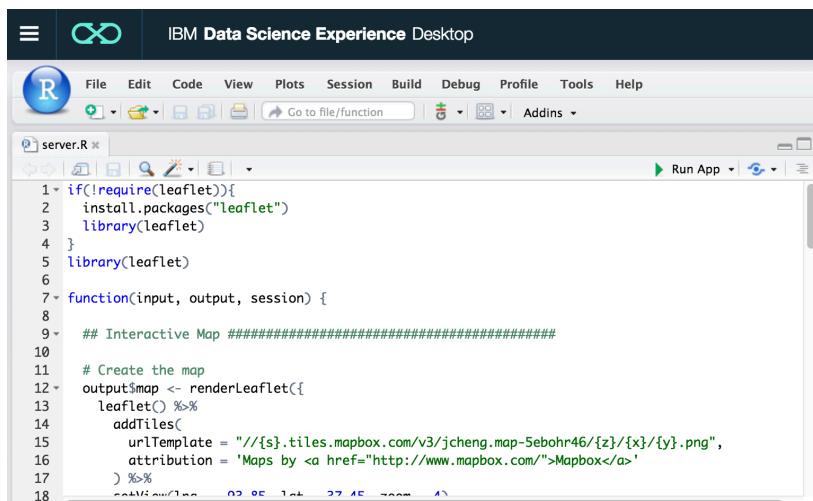


4. Running the flights project:

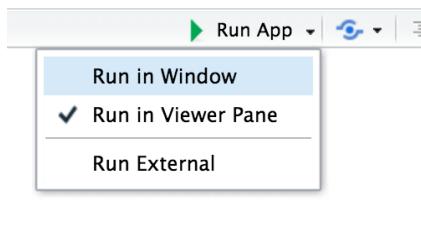
- Click on flights folder



- Click on server.R and it will open in the upper left corner of RStudio



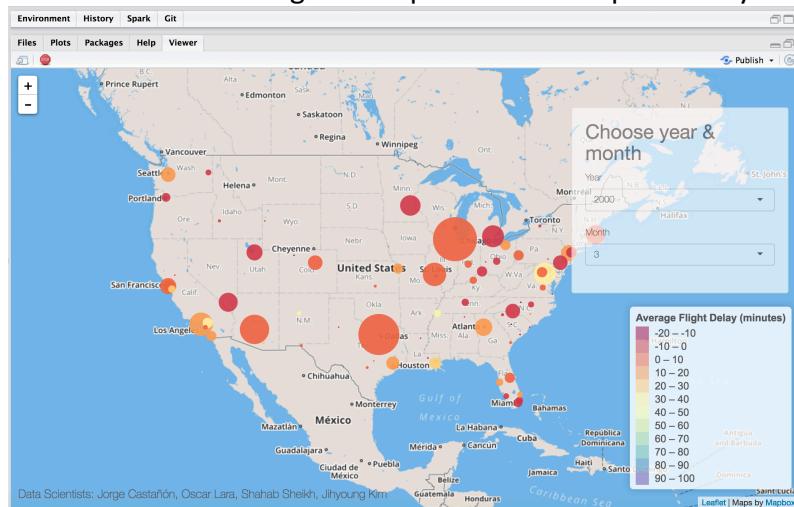
- Click on the arrow next to 'Run App' and select 'Run in Viewer Pane' from the dropdown list



- d. Click on 'Run App'. You should get a cool map on the right bottom pane in the Viewer tab. You can adjust the size of the tab.

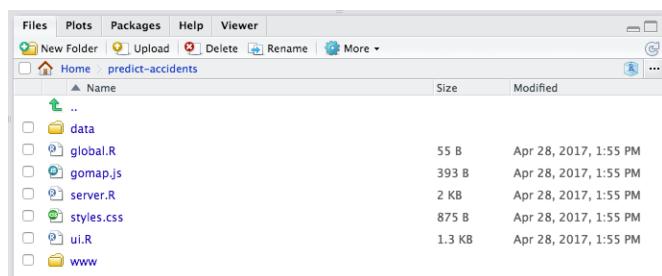
The map shows the probabilities of car accidents per zip code. On the right panel of the app you can interact with the app by choosing the desired date and time of interest.

Click on the circles to get the zip code and the probability.



## 5. Running the predict-accidents project:

- a. Click on the predict-accidents folder



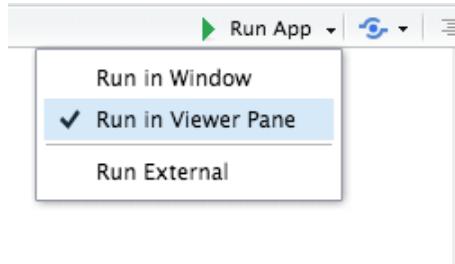
- b. Click on server.R and it will open in the upper left corner of RStudio

```

1+ if(require(leaflet)){
2  install.packages("leaflet")
3  library(leaflet)
4 }
5 library(leaflet)
6
7+ function(input, output, session) {
8
9  ## Interactive Map #####
10
11  # Create the map
12  output$map <- renderLeaflet({
13    leaflet() %>%
14      addTiles(
15        urlTemplate = "//{s}.tiles.mapbox.com/v3/jcheng.map-Sebohr46/{z}/{x}/{y}.png",
16        attribution = 'Maps by <a href="http://www.mapbox.com/">Mapbox</a>'
17      ) %>%
18      addControl(

```

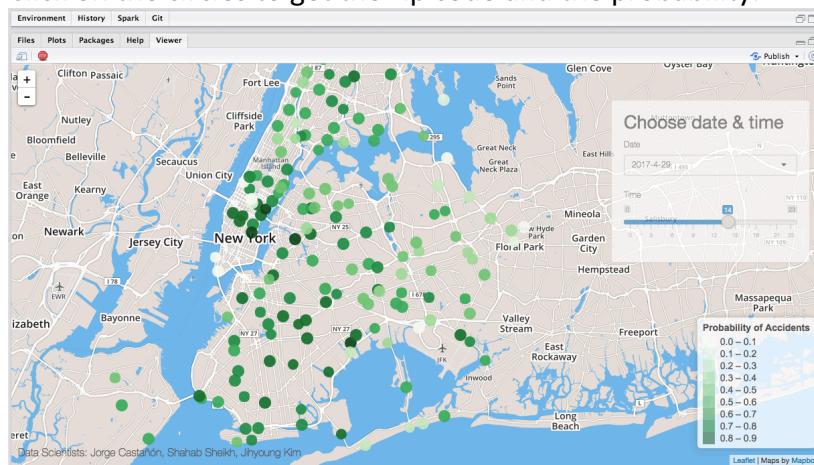
- c. Click on the arrow next to 'Run App' and select 'Run in Viewer Pane' from the dropdown list



- d. Click on 'Run App'. You should get a cool map on the right bottom pane in the Viewer tab. You can adjust the size of the tab.

You should get a cool map that shows the probabilities of car accidents per zip code. On the right panel of the app you can interact with the app by choosing the desired date and time of interest.

Click on the circles to get the zip code and the probability.



## 6. Shiny App – Flex

- a. Download the file from the below link:

<https://raw.githubusercontent.com/IBMDatascience/SparkSummitDemo/master/shinyDemo.Rmd>

- b. Copy the file in the user volume:

For mac:

Copy the folders to location:

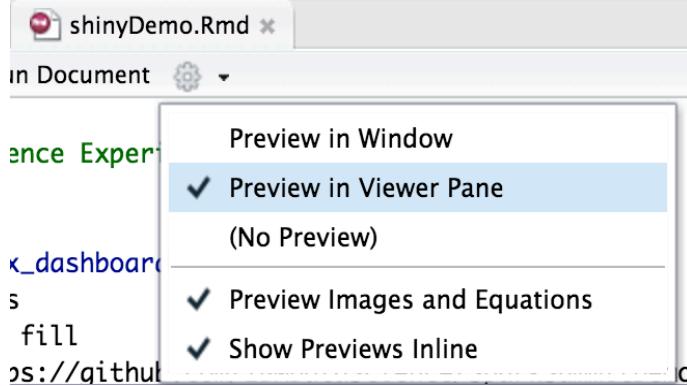
/Users/<username>/Library/Application Support/ibm-dsx-desktop/RStudio/myRFiles/

For Windows:

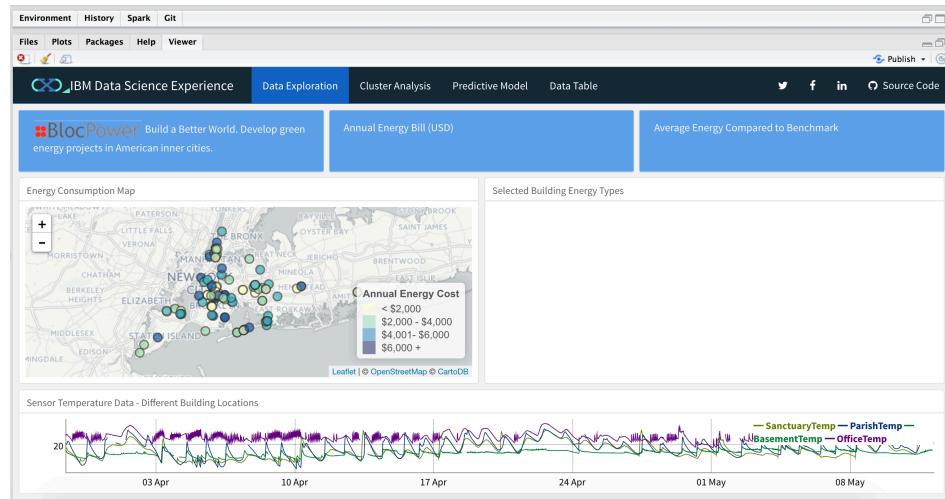
Copy the folders to location:

%APPDATA%\ibm-dsx-desktop\RStudio\myRFiles\

- c. Select lines 21 - 65 and execute (This is a one-time set up to install all necessary packages)
- d. Click on the arrow next to the 'Run Document' and select 'Preview in Viewer Pane'



- e. Click on 'Run Document' and the viewer pane on the right bottom corner will display the results



## Exercise 6: Access IBM Analytics for Apache Spark from RStudio

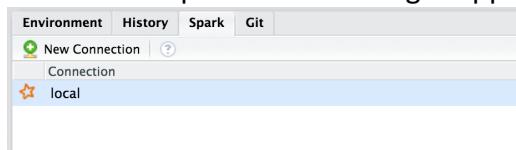
### Tasks:

- Connect to the local spark instance in RStudio
- Loads the popular mtcars R data frame and then generates a Spark data frame for the mtcars data frame. Then do transformations to create a training data set and runs a linear model on the training data set.

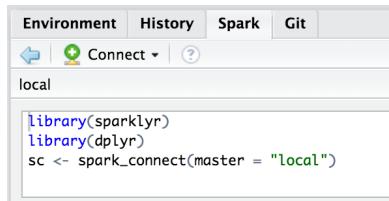
### Description:

#### 1. Connect to local spark

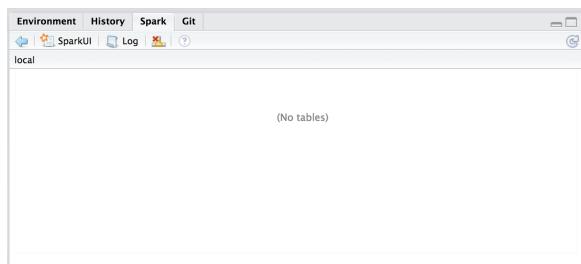
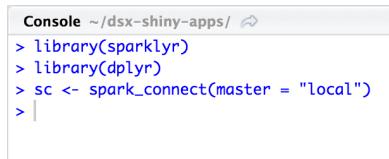
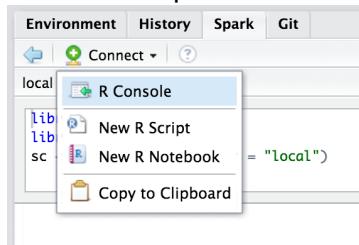
- a. Click on the Spark tab on the right upper tab



- b. Click on local and then click on Connect

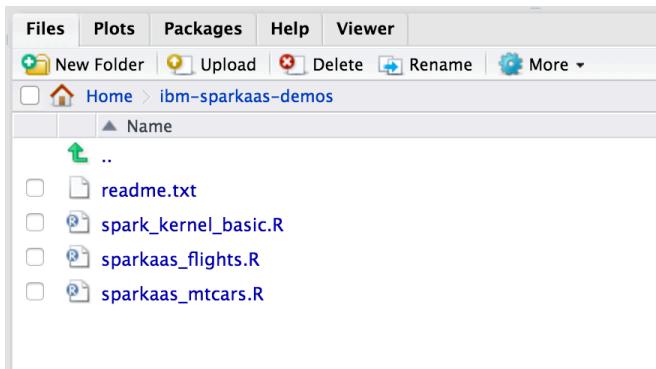


- c. Click on Run on Console. You will see some code lines on the console which will connect to the local spark instance.

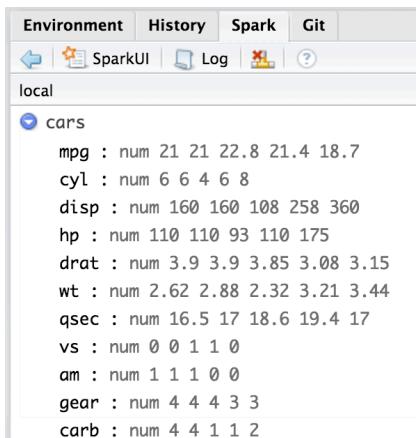


2. Generate a Spark data frame for the mtcars data frame

- a. Click on sparkaas\_mtcars.R and the file opens in the R Script pane



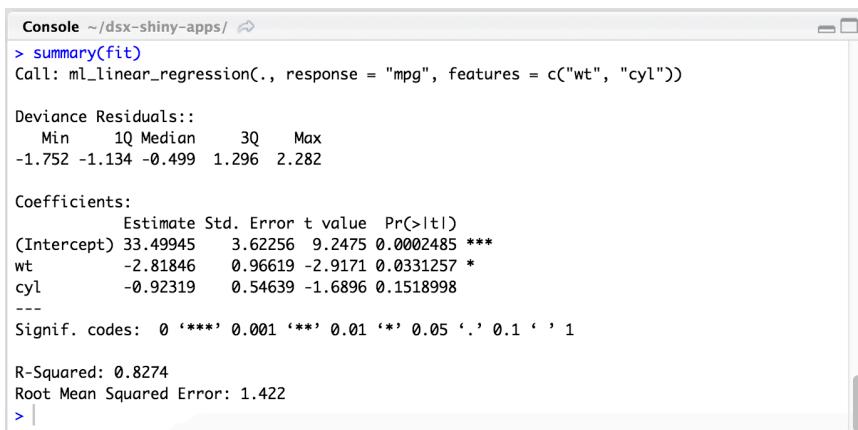
- b. Select all and click on Run.



```
Environment History Spark Git
SparkUI Log

local
cars
  mpg : num 21 21 22.8 21.4 18.7
  cyl : num 6 6 4 6 8
  disp : num 160 160 108 258 360
  hp : num 110 110 93 110 175
  drat : num 3.9 3.9 3.85 3.08 3.15
  wt : num 2.62 2.88 2.32 3.21 3.44
  qsec : num 16.5 17 18.6 19.4 17
  vs : num 0 0 1 1 0
  am : num 1 1 1 0 0
  gear : num 4 4 4 3 3
  carb : num 4 4 1 1 2
```

The results are shown in console as below:



```
Console ~/dsx-shiny-apps/
> summary(fit)
Call: ml_linear_regression(., response = "mpg", features = c("wt", "cyl"))

Deviance Residuals::
  Min    1Q Median    3Q   Max
-1.752 -1.134 -0.499  1.296  2.282

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 33.49945   3.62256  9.2475 0.0002485 ***
wt          -2.81846   0.96619 -2.9171 0.0331257 *
cyl         -0.92319   0.54639 -1.6896 0.1518998
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-Squared: 0.8274
Root Mean Squared Error: 1.422
> |
```