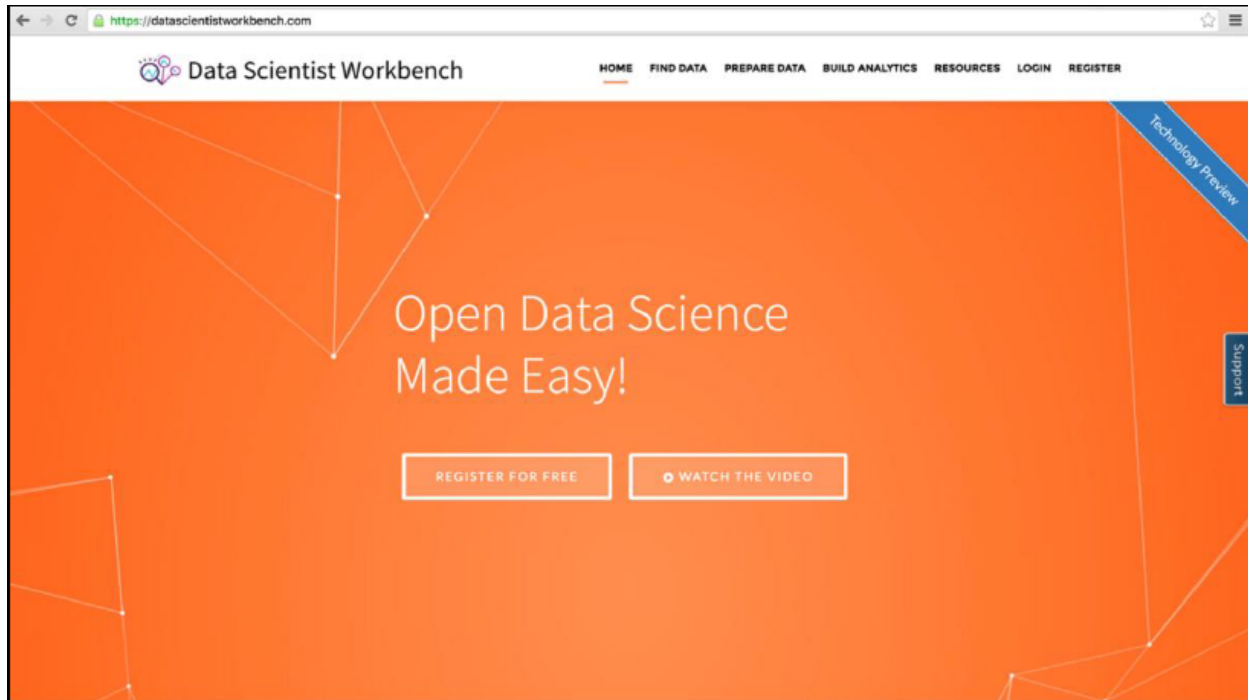




Lab: Introducing Data Scientist Workbench

This Lab must be done on the Data Scientist Workbench. If you have not signed up yet, go to datascientistworkbench.com and click REGISTER FOR FREE. Then login (2nd link from the right):





Contents

LAB 1	INTRODUCING DATA SCIENTIST WORKBENCH	3
	1.1 GETTING STARTED WITH DATA SCIENTIST WORKBENCH	4
	1.2 SUMMARY	4
	1.3 GETTING STARTED WITH DATA SCIENTIST WORKBENCH - SOLUTIONS.....	5



Lab 1 Introducing Data Scientist Workbench

After completing this hands-on lab, you will know:

- How to register an account on Data Scientist Workbench (DSWB).
- How to upload and work with your data on DSWB.
- How to use OpenRefine to prepare your data on DSWB.
- How to open and do simple tasks with the build analytics tools on DSWB, Jupyter & Zeppelin Notebooks, RStudio IDE, and Seahorse.
- How to change your profile settings on DSWB.
- How to open the Feedback Forum and vote on your favourite ideas.
- How to see the courses available for free on Big Data University.

Allow 30 minutes to complete this section of the lab.

NOTE: This is a guided Lab. Solutions are found at the end of this document



1.1 Getting Started with Data Scientist Workbench

- __1. Register an account on DSWB
- __2. Upload data by going to My Data, and Upload Data
- __3. Explore each of the tools: Click OpenRefine
- __4. Click on Jupyter Notebook and explore the interface, like clicking on one of the tutorial notebooks
- __5. Click on Zeppelin Notebook and import one of the sample notebooks
- __6. Click on RStudio IDE and run 1+1 in the console
- __7. Open Seahorse and open one of the tutorials
- __8. Change your profile settings by toggling the left-hand sidebar and clicking on your username (<https://my.datascientistworkbench.com/profile>). Select a preferred data center to optimize your account's performance (upload/download speeds for you)
- __9. Click on the Feedback Forum and read through some of the top ideas. Vote for your favorite idea.
- __10. Click on Online Learning to check out courses on Big Data University

1.2 Summary

Congratulations! ...

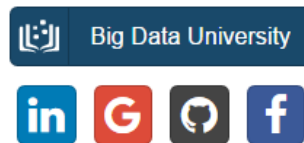


1.3 Getting Started with Data Scientist Workbench - Solutions

1. **Register an account on DSWB:** Go to datascientistworkbench.com and click **SIGN UP FOR FREE**. If you have already registered, simply login. You do not need to do this step. To register, create an account using one of your social media accounts (follow the instructions on the screens that open up for each of the social media, as they differ somewhat):



Create an account using



Or create your account specifying the required information:

First Name*	<input type="text"/>
Last Name*	<input type="text"/>
Email*	<input type="text"/>
Username*	<input type="text"/>
Password*	<input type="password"/>
Confirm Password*	<input type="password"/>
Company	<input type="text"/>

Country	<input type="text" value="United States"/>
Preferred Data Center	<input type="text" value="SoftLayer - Toronto, Canada"/>
Willing to be quoted?	<input checked="" type="radio"/> Yes <input type="radio"/> No <small>If you like what you see, would you be willing to provide a quote or be a reference?</small>
Why are you interested in participating?	<input type="text"/>
Accept License	<input type="checkbox"/> I agree to the Data Scientist Workbench Terms of Service and Honor Code. View License
<input type="button" value="Create your account"/>	
Already have an account? Log in instead!	

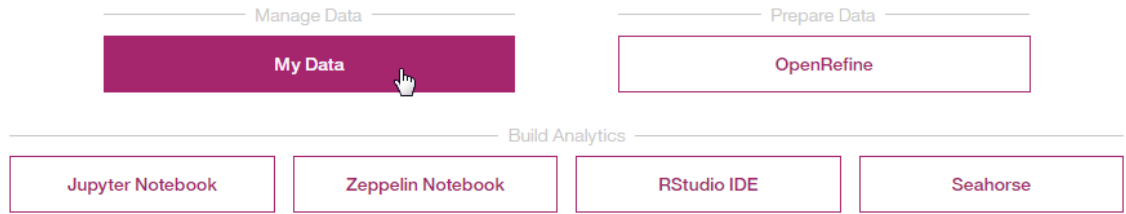
Based on your location, use the default Data Center, or select one from the dropdown.

Then view the License terms, and click Accept License to create your account.

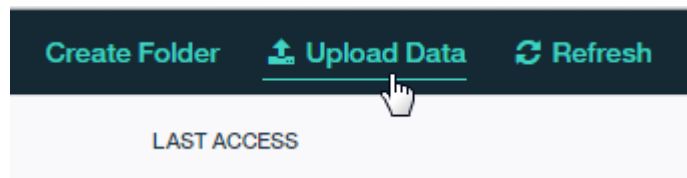


- __2. **Upload data by going to My Data, and Upload Data:** From the DSWB home page, click My Data:

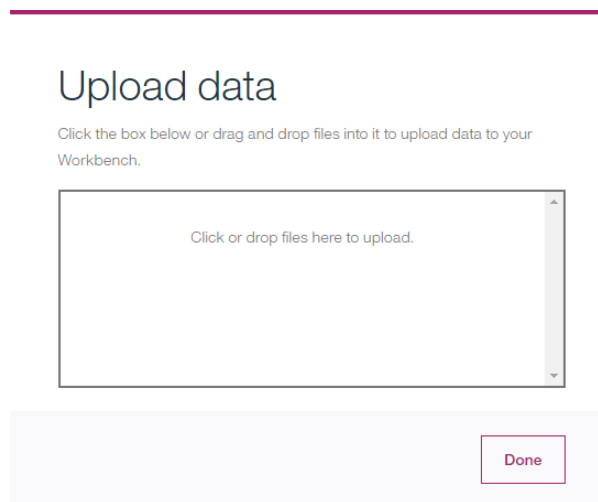
What do you want **to do** today?



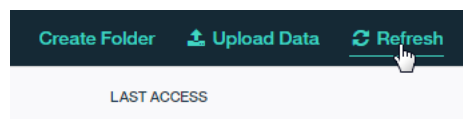
To upload a file to the default `/resources/data` directory, click **Upload Data** (top right):



Drag and drop the files you want to upload and click **Done**:



To check if your files have been uploaded, click **Refresh**:



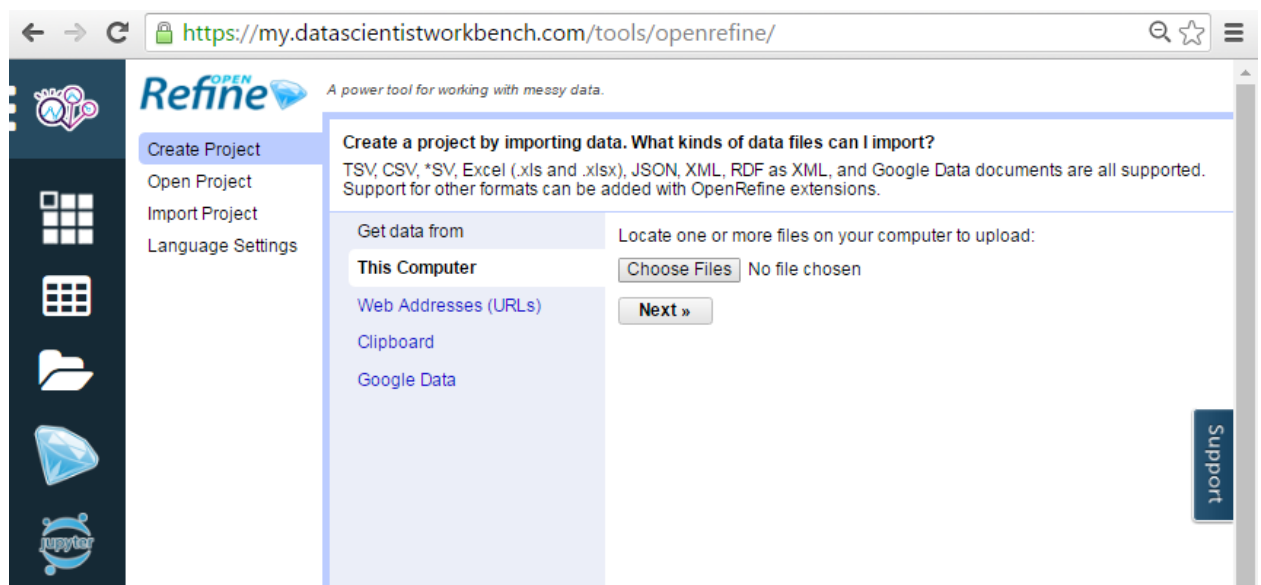


__3. Explore each of the tools: Click OpenRefine:

What do you want **to do** today?

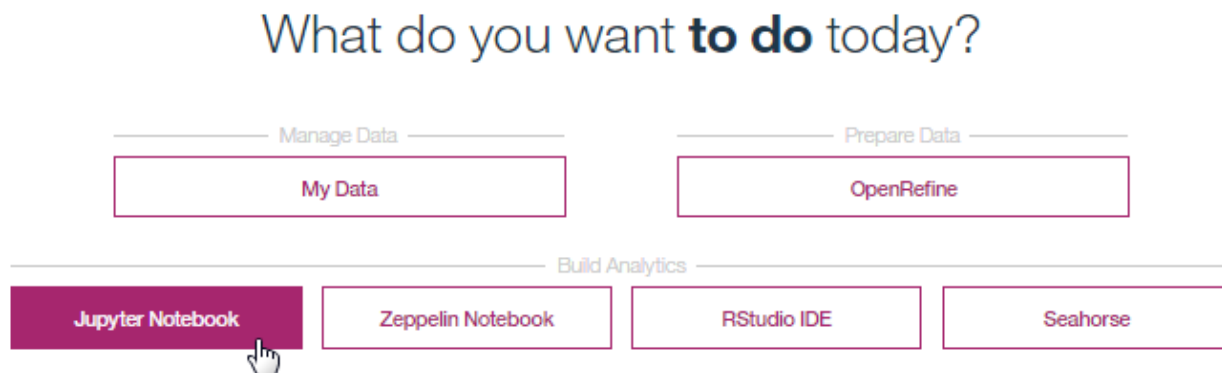


OpenRefine opens – explore some of the options:

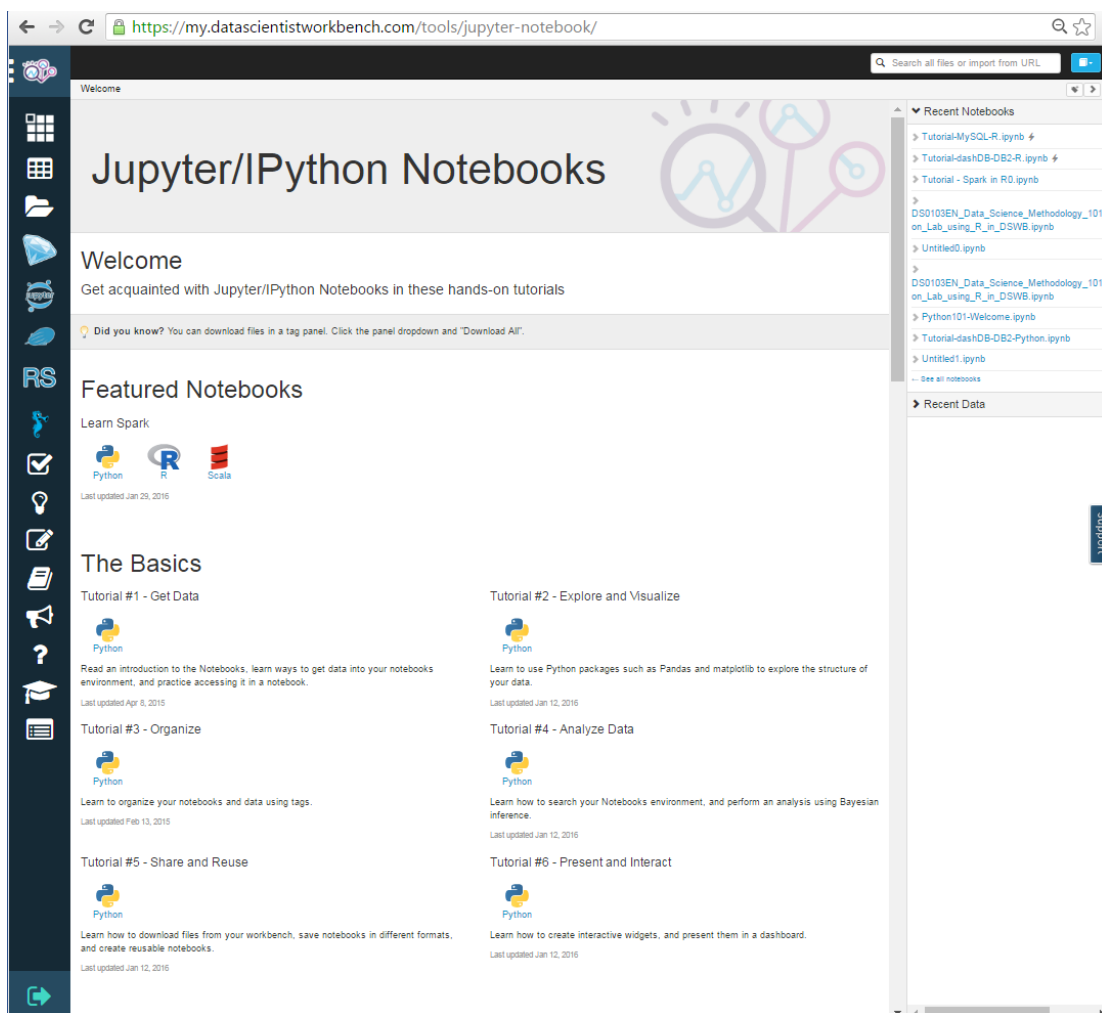




4. Click on Jupyter Notebook and explore the interface, like clicking on one of the tutorial notebooks:



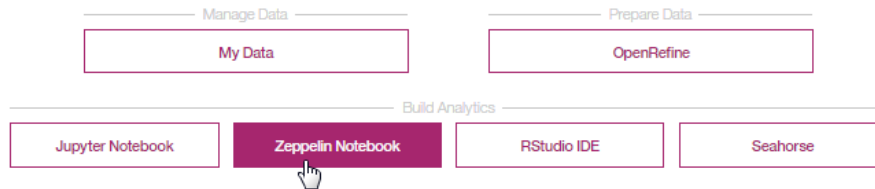
The Jupyter Notebook opens – scroll down and explore some of the tutorials:





5. Click on Zeppelin Notebook and import one of the sample notebooks:

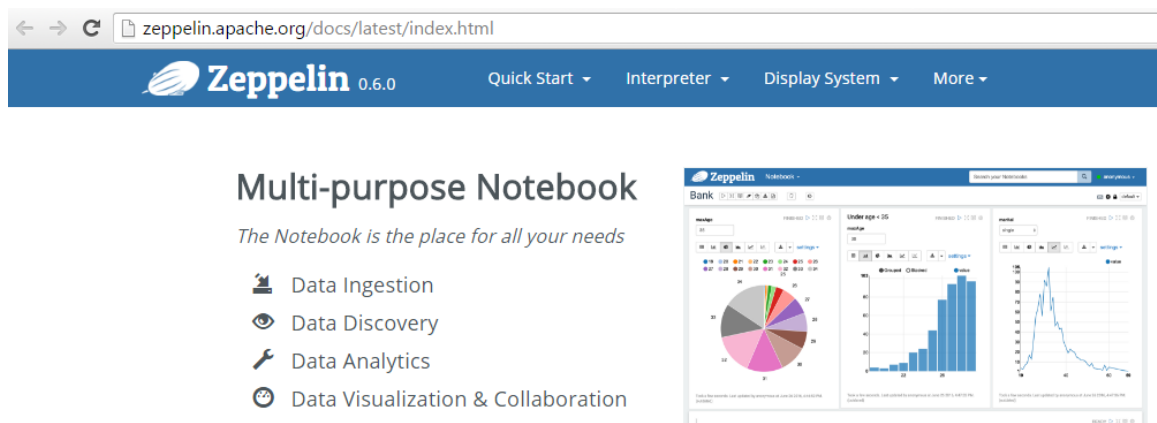
What do you want **to do** today?



The Zeppelin Notebook opens:



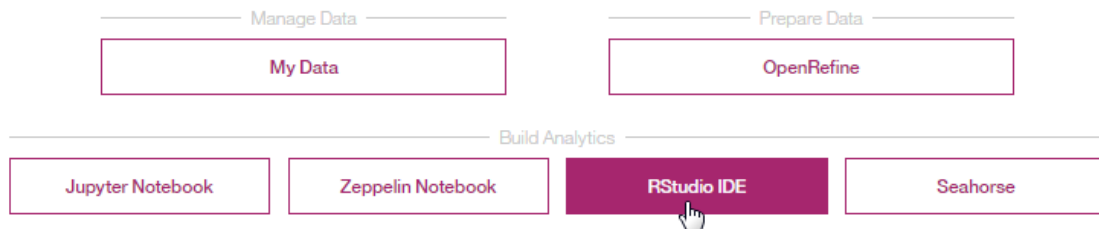
Click Help to explore what this Notebook does:



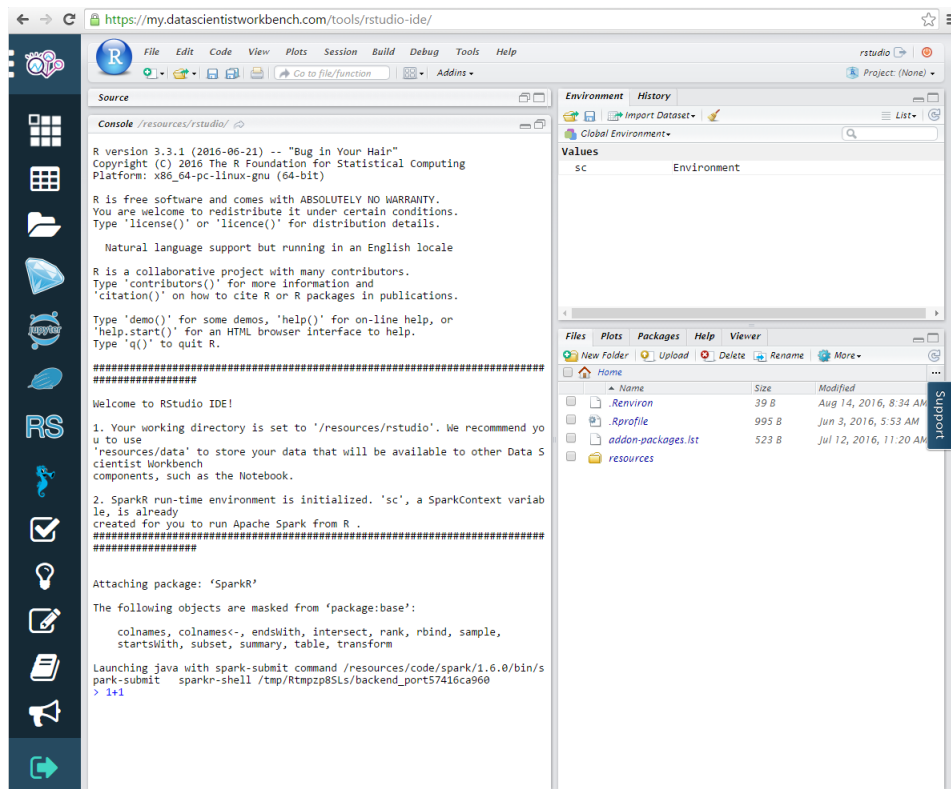


__6. Click on RStudio IDE and run 1+1 in the console:

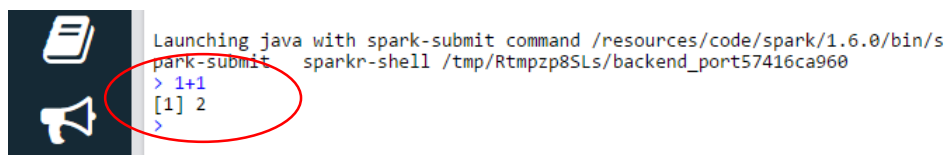
What do you want **to do** today?



RStudio IDE opens; type 1+1 on the Console (left pane) – as shown below; right pane shows environment variables, history, Files, Plots, Packages, Help, and Viewer:



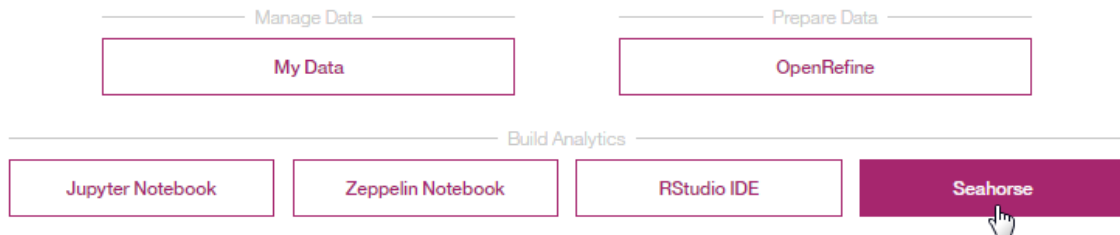
To run the code, press Enter; the output displays below the code:



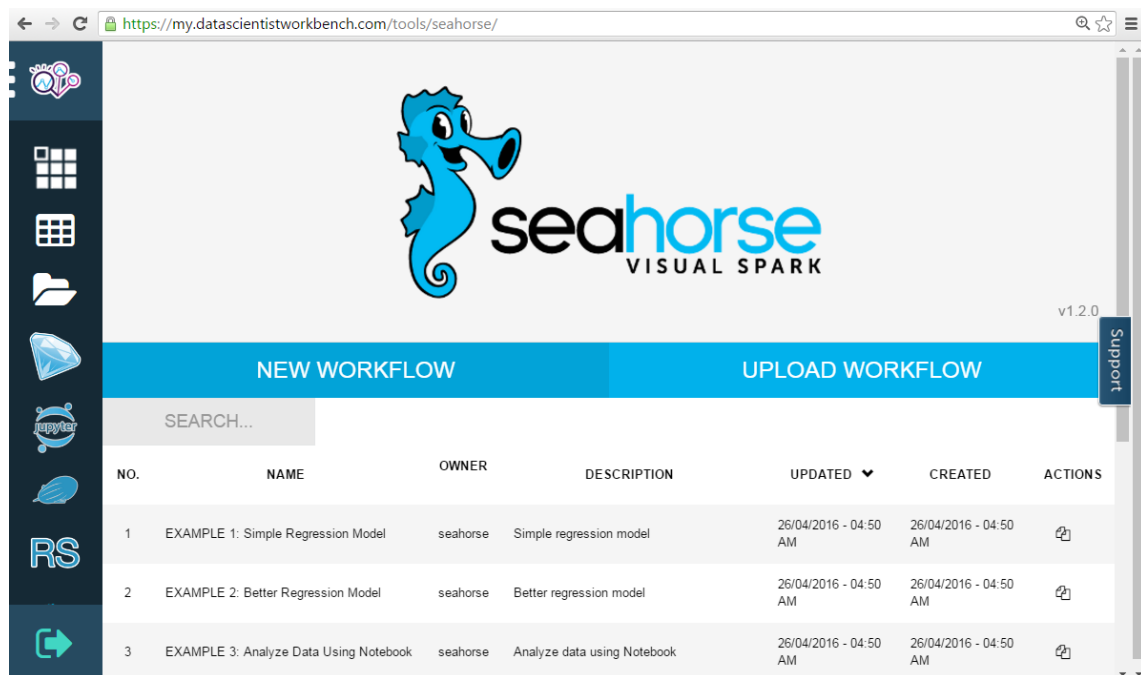


__7. Open Seahorse and open one of the tutorials:




What do you want **to do** today?



Seahorse opens:



Explore one of the tutorials (by clicking on it):

NO.	NAME	OWNER	DESCRIPTION	UPDATED ▼	CREATED	ACTIONS
1	EXAMPLE 1: Simple Regression Model	seahorse	Simple regression model	26/04/2016 - 04:50 AM	26/04/2016 - 04:50 AM	
2	EXAMPLE 2: Better Regression Model	seahorse	Better regression model	26/04/2016 - 04:50 AM	26/04/2016 - 04:50 AM	
3	EXAMPLE 3: Analyze Data Using Notebook	seahorse	Analyze data using Notebook	26/04/2016 - 04:50 AM	26/04/2016 - 04:50 AM	



- ___8. **Change your profile settings** by toggling the left-hand sidebar (bottom left) and clicking on your username:

The screenshot shows the 'My Profile' page in the Data Scientist Workbench. The left sidebar contains the following navigation links:

- FIND DATA**
 - IBM Analytics Exchange
 - Open Data
- MANAGE DATA**
 - My Data
- PREPARE DATA**
 - OpenRefine
- BUILD ANALYTICS**
 - Jupyter Notebook
 - Zeppelin Notebook
 - RStudio IDE
 - Seahorse
- RESOURCES**
 - Workbench Status
 - Submit an idea
 - Blog
 - Knowledge Base
 - Feedback Forum
 - Support
 - Online Learning
 - What's new

The main profile form contains the following fields:

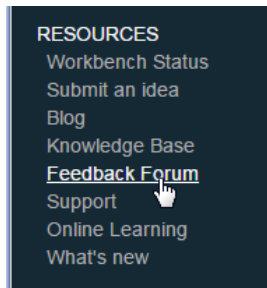
- Profile Picture**: A generic avatar icon. Text: "If you see generic avatar next to your name it means you either don't have a Gravatar for that email address or it is not rated G. Get one for free at: [gravatar.com](\"http://gravatar.com\")"
- First Name**: Big Data
- Last Name**: University
- Email (cannot change)**: bigdatauniversity@gmail.com
- Username (cannot change)**: bigdatauni
- Preferred Language**: English
- Country**: Canada
- Preferred Data Center**: Toronto, Canada - SoftLayer

A **Support** button is located on the right side of the profile form. At the bottom right of the form is an **Update Profile** button.

The footer of the page includes the user's name **bigdatauni** with a **Logout** link, and the text: "Crafted with ❤ by IBM Data Scientist Workbench © Copyright IBM Corp. 2016".



9. Click on the Feedback Forum and read through some of the top ideas. Vote for your favorite idea:



← → ↺ support.datascientistworkbench.com/forums/246775 🔍 ☆ ☰

Data Scientist Workbench

Workbench
← IDEAS? What Should We Do Next?

I suggest you ...
Enter your idea

Hot ideas Top New Category Status My feedback

4 votes
Vote

create folder not working
The "create folder" is not working. I've tried both Chrome and IE, neither works.
3 comments · Data Prep · Flag idea as inappropriate...

7 votes
Vote

Add support for GitHub into RStudio
Is it possible to add git hub support to do versioning control with git hub as we do in the Desktop version of RStudio? I would love to fork some projects and do commits using RStudio on dswb.
0 comments · Rstudio IDE · Flag idea as inappropriate...

3 votes
Vote

Allow direct linking to an existing notebook
A good enhancement would be to allow direct linking to an existing notebook. Instead of automatically creating a duplicate, with a suffix, why not just open the existing notebook.
0 comments · General · Flag idea as inappropriate...

4 votes
Vote

Allow folder synchronization with Google Drive
It would be great if we could selectively synchronize folders with google drive / dropbox as it would allow us to automatically backup the code as well as the data. In addition it would be possible to synchronize these with local devices which in turn can allow us to work on the code in local machines which are running in areas with poor internet connectivity.
0 comments · General · Flag idea as inappropriate...

31 votes
Vote

Integrate Airbnb's Caravel project as Data Exploration tool
Add Caravel new Open Source project from Airbnb in the DSWB. See a quick demo here: https://www.youtube.com/watch?v=3Txm_nj_R7M
Caravel's main goal is to make it easy to slice, dice and visualize data. It empowers users to perform analytics at the speed of thought.

Agatha Colangelo (Stay signed in)
Settings Sign out

Workbench
Post a new idea...

All ideas
My feedback
Data Prep (7)
General (60)
Jupyter - Python (5)
Jupyter - R (12)
Jupyter - Scala (3)
Jupyter Notebooks (38)
Modeler (1)
Rstudio IDE (3)
Spark (4)

Search

Contact support

Give feedback
Data Preparation (1)
Jupyter Notebooks (16)
Rstudio IDE (4)
Workbench (138)
Zeppelin Notebooks (0)

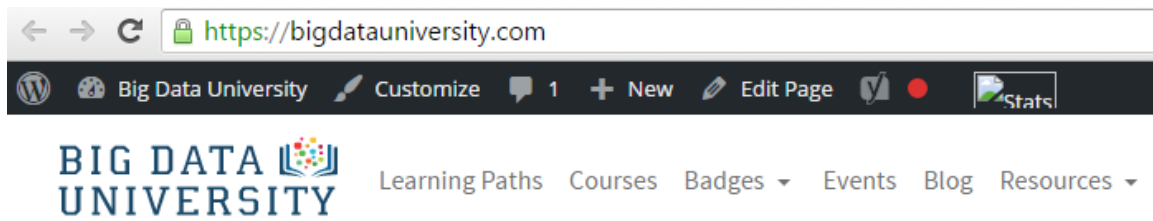
Knowledge Base
General (8)
Jupyter Notebooks (21)
OpenRefine (1)
RStudio IDE (2)
Zeppelin Notebooks (1)
All articles

IBM | Data Scientist Workbench

Lab – Introducing Data Scientist Workbench Page | 13



__10. Click on Online Learning to check out courses on Big Data University:



Check out the Learning paths:

Learning Paths
Pick one of our selected learning paths, and get started today!

FEATURED [View All](#)

Learning Path	Badges	Courses	Description
Big Data Fundamentals	4	3	Are you interested in understanding 'Big Data' beyond the terms used in headlines? Then select this learning path as an introduction to tools like Apache Hadoop and Apache Spark Frameworks, which enable data to be analyzed on mass, and start the journey towards your headline discovery.
Scala Programming for Data Science	2	3	To light a fire, do you use a match, a lighter, or a torch? Depends on the size of the fire, much like the decisions that lead one to use Python, R, or Scala. Spark your interest in selecting the tools you need to tackle Big Data with ease, that will not just blow out.
Hadoop Fundamentals	5	4	Are you interested in moving beyond the elephant in the room and understanding Hadoop as a foundational tool set in your future? Then select this learning path to gain exposure to the tools used in Big Data, Hadoop's core components and supporting open source projects.
Spark Fundamentals	2	3	Solid understanding and experience, with core tools, in any field promotes excellence and innovation. Apache Spark, as a general engine for large scale data processing, is such a tool within the big data realm. This learning path addresses the fundamentals of this program's design and its application in the everyday.
Data Science for Business	1	2	If a trend is identified using big data, how can it be applied to solving complex day-to-day problems? Are there implications of using this data? This learning path addresses the principles of data science to explore data privacy, regression analysis, text analytics, data visualization and predictive modeling to promote topical awareness.
Big Data Analytics	2	3	Is the thought of writing code limiting your interest to visualize and analyzing big data efficiently? If so, the Big Data Analytics learning path offers a foundation of both methods and tools to go beyond coding and equip you to achieve the results you require using Big Data.



















And some of the free courses:

← → ↻ <https://bigdatauniversity.com/courses/> 🔍 ☆ ☰

BIG DATA UNIVERSITY Learning Paths Courses Badges ▾ Events Blog Resources ▾ 🔍 Explore new learning opportunities agathacolangelo ▾

COURSES

 <p>Big Data 101 Big Data University BD0101EN Beginner</p>	 <p>Hadoop 101 Big Data University BD0111EN Beginner</p>	 <p>Spark Fundamentals I Big Data University BD0211EN Beginner</p>	 <p>Machine Learning 101 Big Data University ML0101EN Beginner</p>
 <p>Watson Analytics 101 Big Data University WA0101EN Beginner</p>	 <p>Deep Learning 101 DeepLearningTV ML0115EN Intermediate</p>	 <p>Scala 101 LightBend SC0101EN Beginner</p>	 <p>Data Journalism: First Steps, Skills and Tools Learn2.net DJ0101EN Beginner</p>
 <p>Predictive Modeling Fundamentals I Big Data University PA0101EN Intermediate</p>	 <p>Data Science Methodology Big Data University DS0103EN Beginner</p>	 <p>R 101 DataCamp RP0151EN Beginner</p>	 <p>Analyzing Big Data in R using Apache Spark Big Data University RP0105EN Beginner</p>
 <p>Spark Overview for Scala Analytics LightBend SC0103EN Beginner</p>	 <p>MapReduce and YARN Big Data University BD0115EN Intermediate</p>	 <p>Apache Pig 101 Big Data University BD0121EN Beginner</p>	 <p>Simplifying Data Pipelines with Apache Kafka Big Data University BD0123EN Intermediate</p>

Support