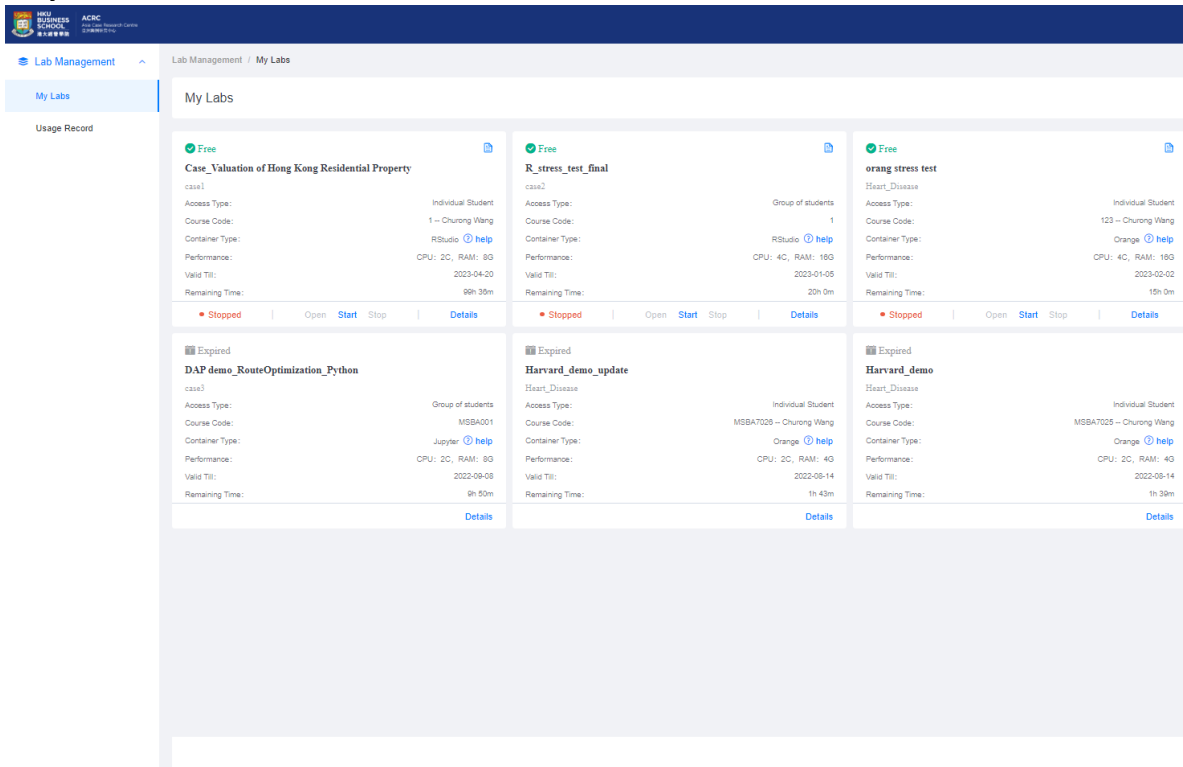


# Data Analytics Platform Operation Manual

## Student

After lecturers subscribe to a case, students will receive a notification email including information on the link to the Data Analytics Platform, a username (email address of the user), and a temporary password. Students could click on the website link <https://dap.acrc.hku.hk/hku-dap-client/#/Signin> to enter the Data Analytics Platform. After entering the email addresses and temporary password, students will be notified to change the temporary password to their password (8-20 digits). Log in to the Platform with the user's email address and password.

## My Labs



Screenshot of the user interface of the Data Analytics Platform



Screenshot of the Labs that students are going to use

## Instructions for using Rstudio

'My Labs' show all the labs that teachers subscribed to for students.

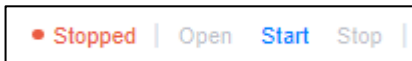
1. For detailed instructions on using the virtual environments, click 'help'.



2. For instructions on the case, click the upper right icon.



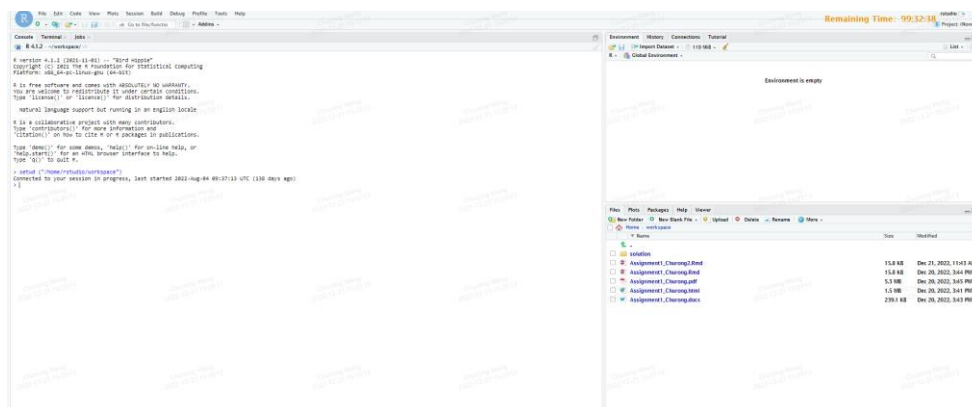
3. To start, open or stop the labs, click the buttons at the bottom of each lab. The left bottom icon shows the status of the lab. Wait for around 15 seconds for the lab to load.



When the status icon turns to "Running," click "Open" to access the virtual environment of Rstudio.



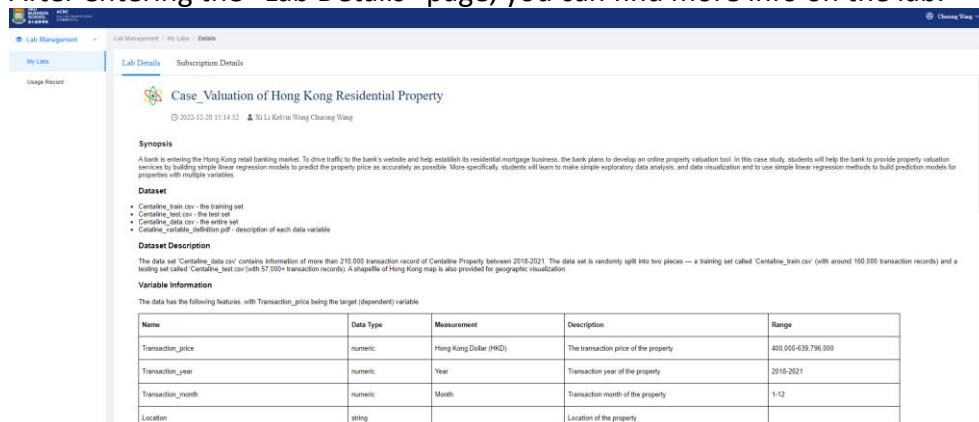
The system will then automatically jump to the virtual environment of Rstudio's user interface. "Remaining Time" shows the time limit of your lab usage. If you use up the lab time, you will need to email the lecturer or the teaching assistant to renew the lab. The watermark shows the user name and the access time of users.



4. Students can view detailed information about each lab by clicking 'Details.'



After entering the "Lab Details" page, you can find more info on the lab.

A screenshot of the "Lab Details" page for a case titled "Case Valuation of Hong Kong Residential Property". The page includes a synopsis, dataset information, and a table of variable information.

**Synopsis**  
A bank is entering the Hong Kong retail banking market. To drive traffic to the bank's website and help establish its residential mortgage business, the bank plans to develop an online property valuation tool. In this case study, students will help the bank to provide property valuation services by building simple linear regression models to predict the property price as accurately as possible. More specifically, students will learn to make simple exploratory data analysis, and data visualization and to use simple linear regression methods to build prediction models for properties with multiple variables.

**Dataset**

- `Casestate_train.csv` - the training set
- `Casestate_test.csv` - the test set
- `Casestate_data.csv` - the entire set
- `Casestate_variable_definitions.pdf` - description of each data variable

**Dataset Description**  
The data set `Casestate_data.csv` contains information of more than 210,000 transaction record of Casestate Property between 2018-2021. The data set is randomly split into two pieces — a training set called `Casestate_train.csv` (with around 160,000 transaction records) and a testing set called `Casestate_test.csv` (with 57,000+ transaction records). A shapefile of Hong Kong map is also provided for geographic visualization.

**Variable Information**  
The data has the following features, with `Transaction_price` being the target (dependent) variable.

Name	Data Type	Measurement	Description	Range
<code>Transaction_price</code>	numeric	Hong Kong Dollar (HKD)	The transaction price of the property	400,000-630,796,000
<code>Transaction_year</code>	numeric	Year	Transaction year of the property	2018-2021
<code>Transaction_month</code>	numeric	Month	Transaction month of the property	1-12
<code>Location</code>	string		Location of the property	

## Dataset Access

The system creates the dataset folder by default, and all datasets are stored under the dataset folder. The following codes are examples to access the datasets:

```
1. ## Set working directory
2. setwd(dir= '/dataset')
3. dir()
```

```
[1] "Centaline_data" "Centaline_test" "Centaline_train" "HKG_adm1"
```

You can see there are some datasets stored under the directory “/dataset”. Select your desired dataset from the directory:

```
1. ## load dataset
2. df = read.csv('/dataset/Centaline_data/Centaline_data.csv', header
  = TRUE)
3. head(df,10)
```

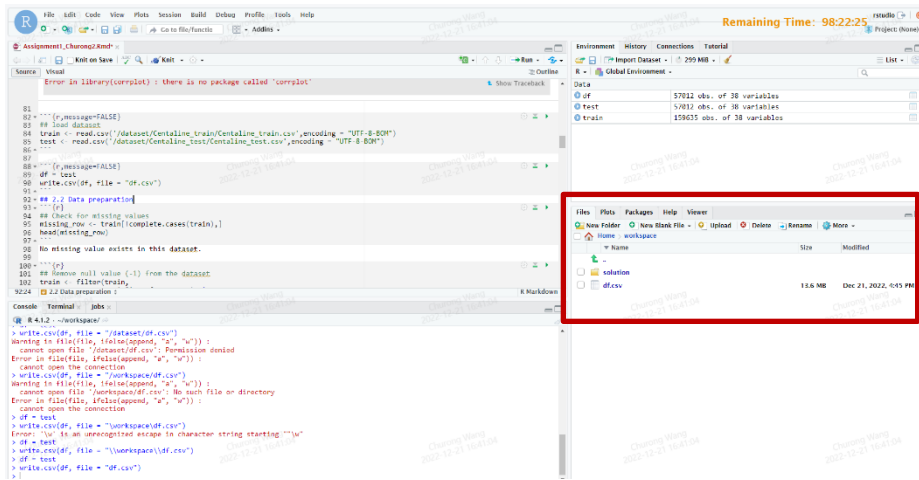
	Transaction_price	Transaction_year	Transaction_month	Location
1	2000000	2021	11	6 SHUN PING STREET
2	3000000	2020	2	200 SHA TAU KOK ROAD SHEK CHUNG AU
3	9300000	2020	3	31 SHUN LUNG STREET
4	9200000	2020	3	31 SHUN LUNG STREET
5	6205200	2020	1	31 SHUN LUNG STREET
6	6170900	2020	8	31 SHUN LUNG STREET
7	6049800	2020	9	31 SHUN LUNG STREET
8	5871900	2020	10	31 SHUN LUNG STREET
9	5800000	2020	1	31 SHUN LUNG STREET
10	5711000	2021	4	31 SHUN LUNG STREET

1 - 10 of 10 rows | 1 - 5 of 38 columns

## Dataset Saving

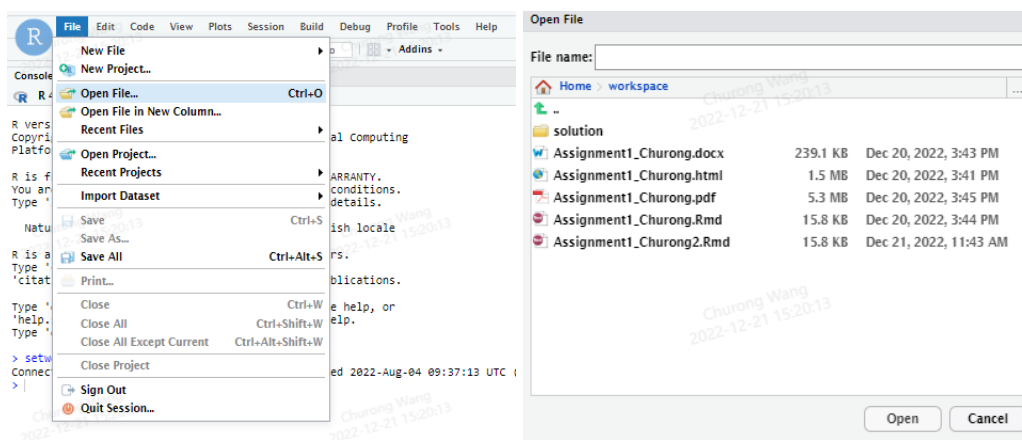
To save a newly created dataset as a CSV file, use the following sample codes to save it in the ‘workspace.’ Then, you may find your dataset under “workspace” in the bottom right section of the user interface.

```
1. ## Save dataset to workspace
2. write.csv(df, file = "df.csv")
```



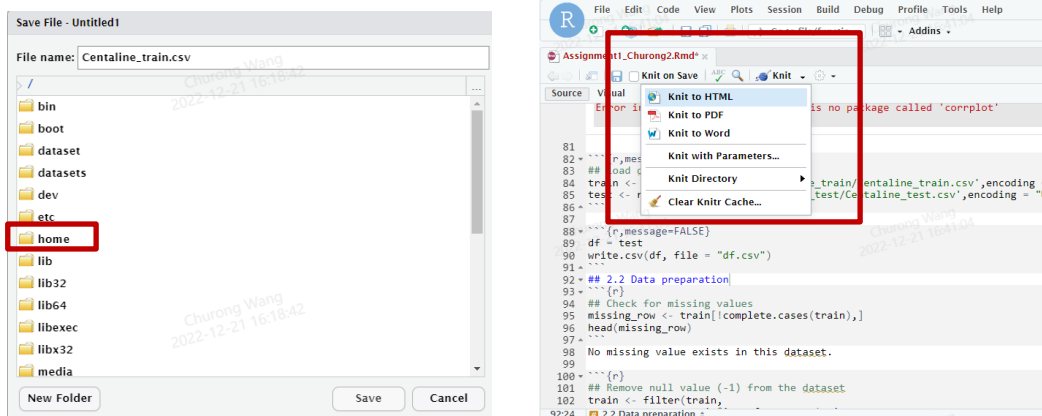
## File Access

All files are stored in a folder called “workspace.” Access the “workspace” by clicking “File > Open File > workspace.”

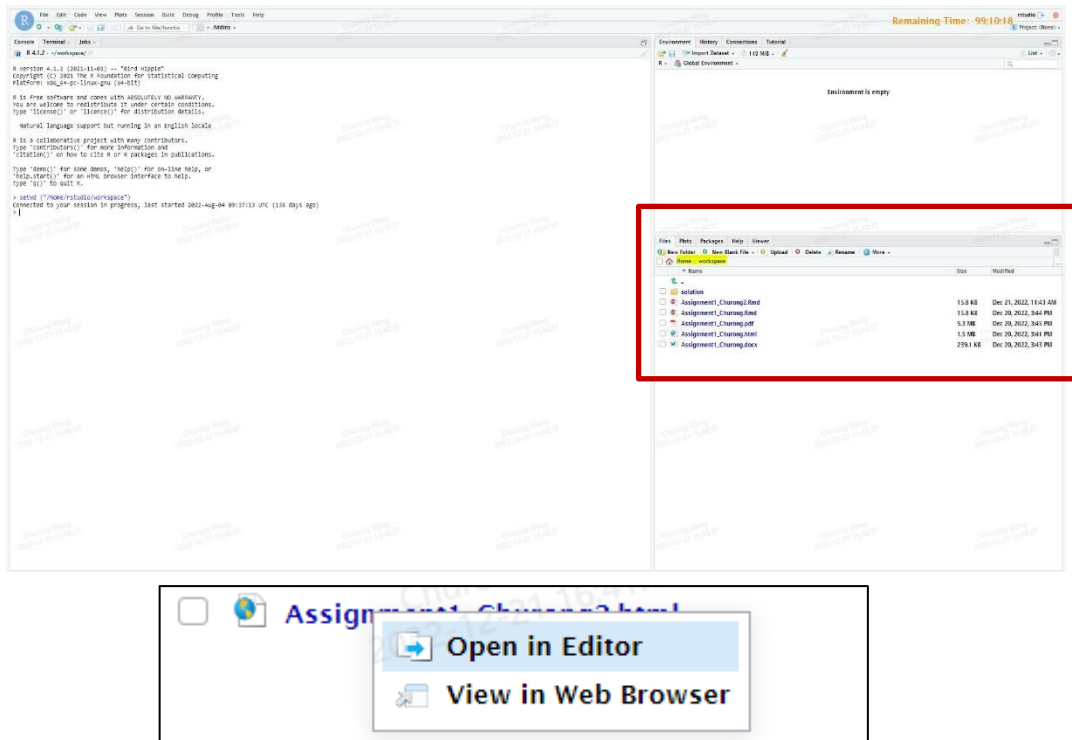


## File Saving

All files must be saved in the “workspace” for future access. Save R files to “workspace” by clicking “File > Save” or “File > Save As...” and browse to the directory “/ > home > rstudio > workspace.” Export R codes to HTML/PDF/Word format by clicking ‘Knit > Knit to ...’.



To check files other than R format (e.g., PDF/HTML/Word format), check out files on the bottom right of the user interface. The “Files” section displays all the files under the folder ‘workspace.’ Double-click the files to see them on a pop-up web page. For HTML format files, click it and select “View in Web Browser” to open the HTML file in a new web page.



### Final Note:

- 1. Please save all files into the folder “workspace,” and ensure that your UID (student ID) and name are contained in the file name when saving your files. These two steps are very important because if you don't follow these instructions, your assignment may not be visible/identifiable to course instructors to grade your assignment.**
- 2. Labs will automatically close after 30 mins of inactivity. Please always remember to save ALL your working files to “workspace” so you can retrieve your previous work in the future.**
- 3. To save your lab time, remember to click ‘Stop’ to close the labs when you do not use the lab service.**
- 4. For group access, please note that the platform does not support simultaneous editing. It is recommended that each group member create a separate file for editing and merge the files for review or submission.**