A quick introductory run-through of Domino

Table of Contents

Introduction

This document will guide you through the process of familiarizing yourself with Domino's interface and basic feature set. It will cover starting a new project, working with interactive notebooks, editing files, retrieving previous versions of files, and launching and comparing runs. It is not intended as a comprehensive training document.

If you have not already done so, the first step is to create an account on Domino. Open a new Chrome window, go to https://app.dominodatalab.com/signup (or contact your administrator if you have an on-premises or VPC installation), and fill out the form on that page to create a new account.

WARNING: On-premises or VPC users of Domino should retrieve the correct URL to use from support@dominodatalab.com or from your administrator. If you do not use the appropriate URL, you will not be able to collaborate with other team members.

You may also wish to download the Domino Command Line Interface (CLI) . To do so, start by clicking the blue button that reads "Download the client for [your operating system]."

Once the CLI package downloads, open it and follow the prompts to install it.

To find out more about how to use the CLI, see the official documentation on our Support site: http://support.dominodatalab.com/hc/en-us/articles/204842905

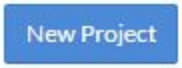
Start a new project

We will now start a new project using the set of crime maps data you have been provided.

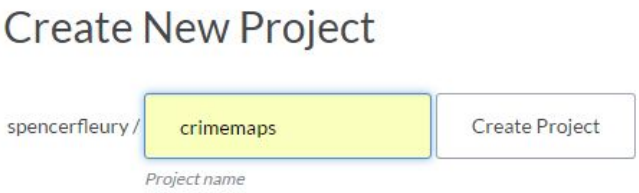
Begin by clicking the `Skip Tour` button. This cancels the guided tour of Domino's features and functionality so you can begin working immediately. If you like, you will be able to return to the tour by clicking the `Tour` button the top of the page.

You should now be looking at the `Projects Overview` page. A Domino project contains code, data, results--generally, all the files and data that make up a data science project.

From the `Projects Overview` page, click the `New Project` button.

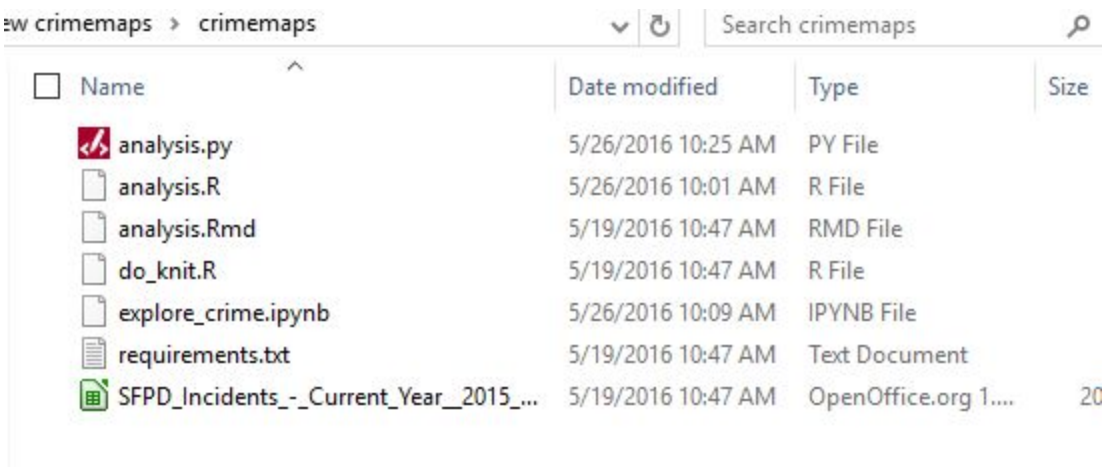


You will be asked to name this project. You should give your project the same name as the corresponding folder on your local machine, especially if you plan to use the CLI. Since the ZIP file that you will download will create a `crimemaps` folder, let's name this project `crimemaps`. Type the name into the text box and click the `Create Project` button.



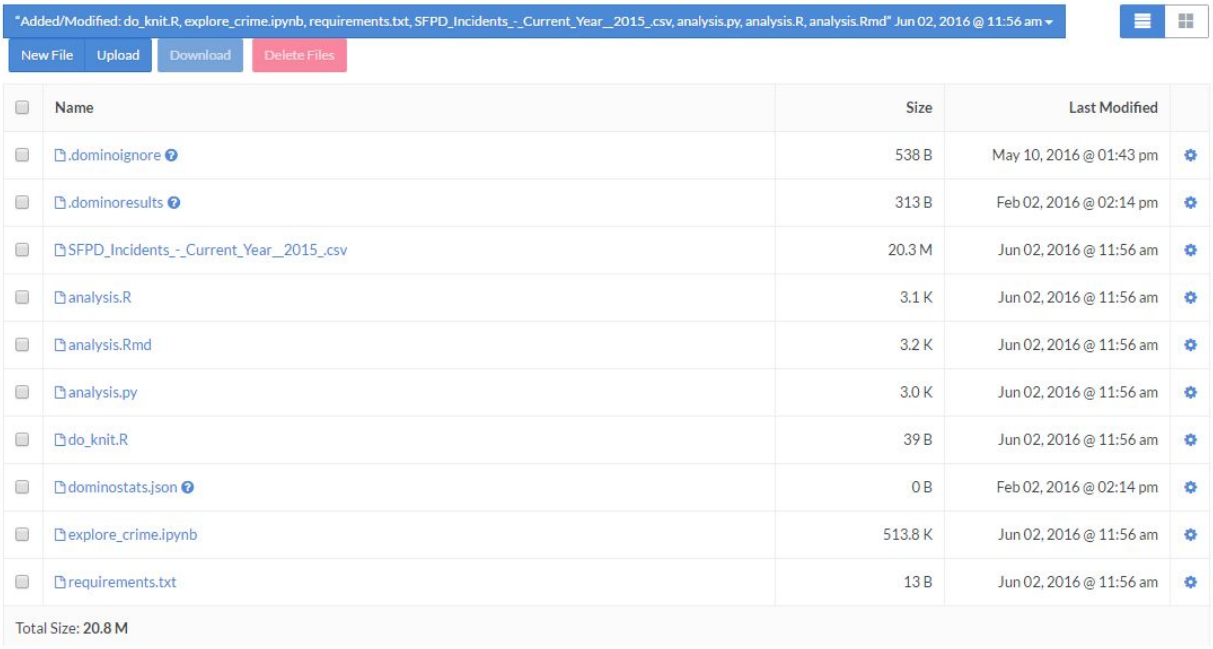
After Domino creates your project, you should be looking at the `Files` tab. You will notice that Domino has already created a few files and added them to this project. These files are not enough to conduct an analysis on their own, so you will have to upload files to this project.

First, create a directory called `crimemaps` somewhere on your computer. Then unzip `crimemaps.zip` into that directory.



Finally, drag the individual files into the UI uploader (***note***: do NOT drag your crimemaps directory into the uploader).

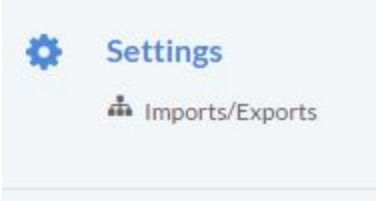
Refresh the page. Your files directory should look like this:



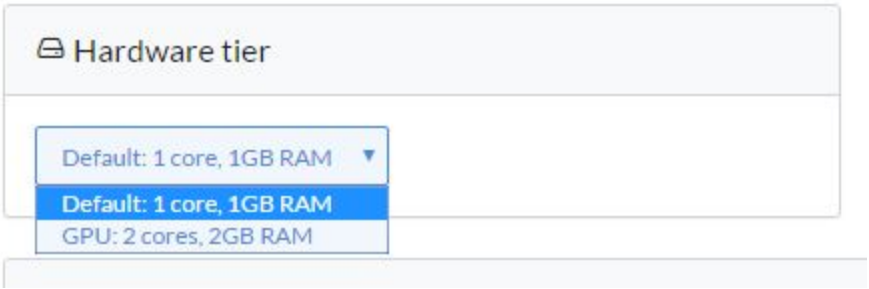
Selecting your hardware tier

Domino allows you to specify the most appropriate hardware tier for the analysis you wish to run.

In the left-hand sidebar, click `Settings`.



This opens up the `Project settings` page. To select a hardware tier, first click on the `Hardware & Environments` tab, if it is not already selected. Then click the drop-down list box in the `Hardware tier` section.

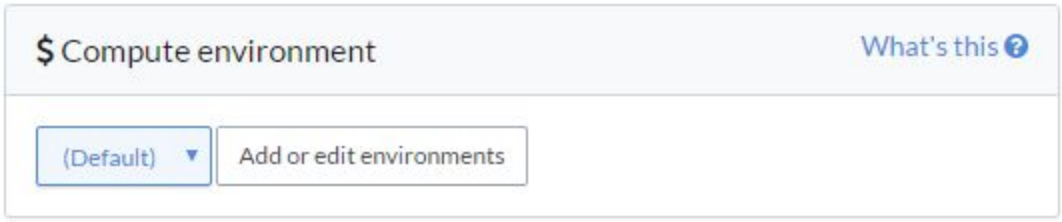


The various hardware tiers available to you are listed here. If you do not see the tier you need, contact Domino support for assistance.

Setting your compute environment

When Domino starts a run, it creates a Domino container based on the "environment" associated with your project. In certain circumstances--usually involving the installation of custom software packages--you may wish to customize your compute environment to accommodate your organization's specific needs.

You can set your compute environment from the `Project settings` page as well. On the `Hardware & Environment` tab is a section labeled `Compute environment`.



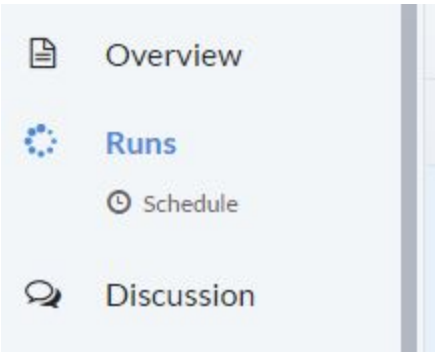
At this point, the only environment available to you should be the default environment. To add a new compute environment, click the `Add or edit environments` button.

By default, user accounts do not have permission to modify environments. Contact Domino to find out how this restriction can be lifted. For more information about environments in Domino, see the official documentation: http://support.dominodatalab.com/hc/en-us/articles/206172876.

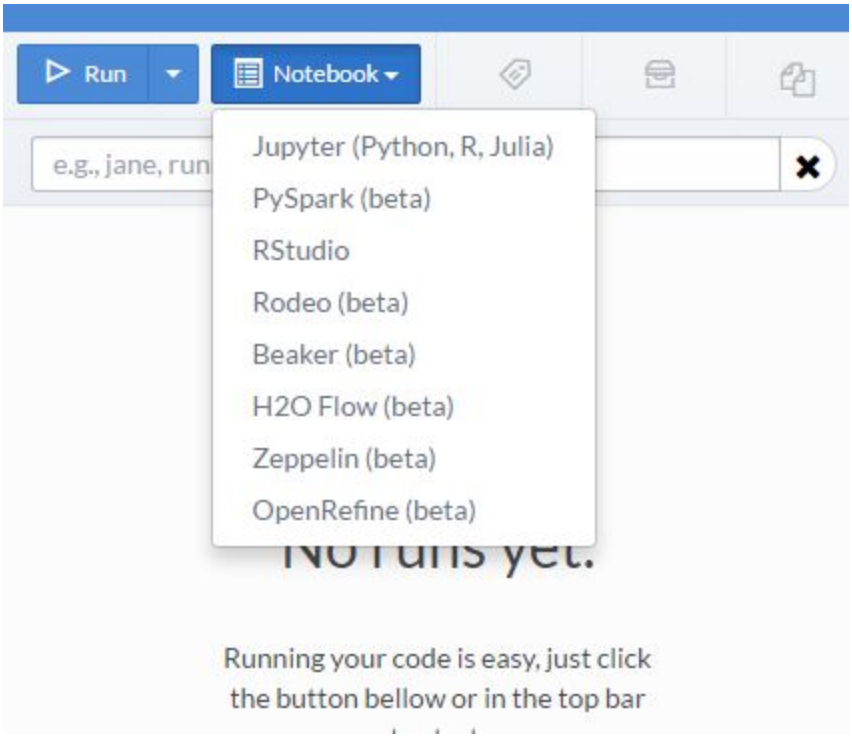
Work in an interactive notebook session

Interactive notebooks are environments that allow for exploratory data analysis, allowing users to easily and quickly experiment with new ideas or data sets. Domino supports several different types of interactive notebooks, including Jupyter and Rstudio.

When Domino starts a machine to host an interactive session, run a script, or anything else, Domino calls that act a *run*. These are controlled from the `Runs` tab. Navigate to the `Runs` tab by clicking on `Runs` in the left-hand sidebar.



In the screen's middle panel, click the `Notebook` button to see a list of interactive notebooks you can use for your Domino session.

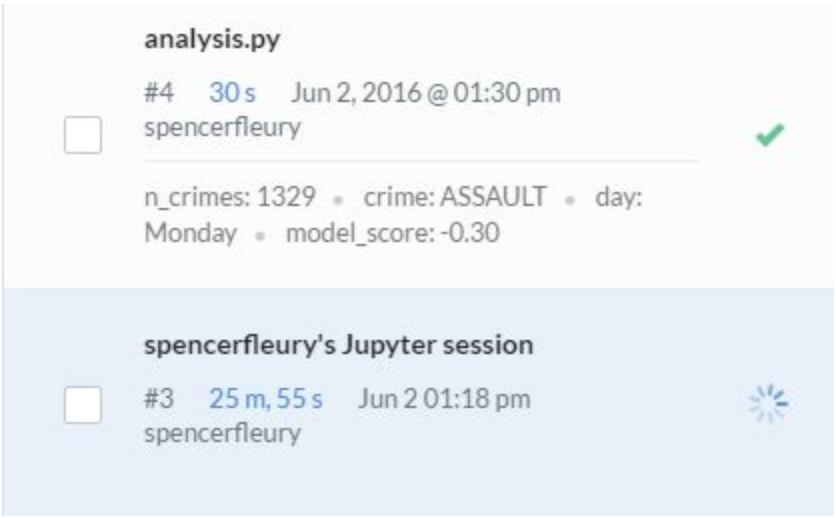


For Python users

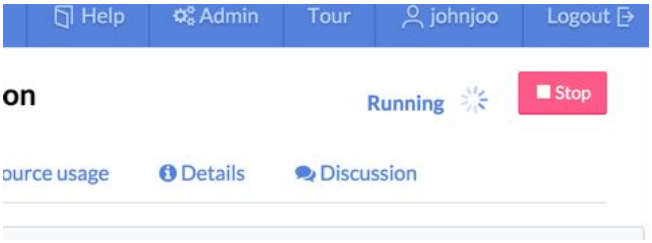
Select Jupyter from the drop-down list. In the right-hand panel, click the `Open Session` button. A Jupyter session will launch.

Click on `explore_crime.ipynb` to open and explore the file. Run the file to generate some initial results.

When you are done with the run, you will have to stop the session. Select the run for the Jupyter session that is currently active.



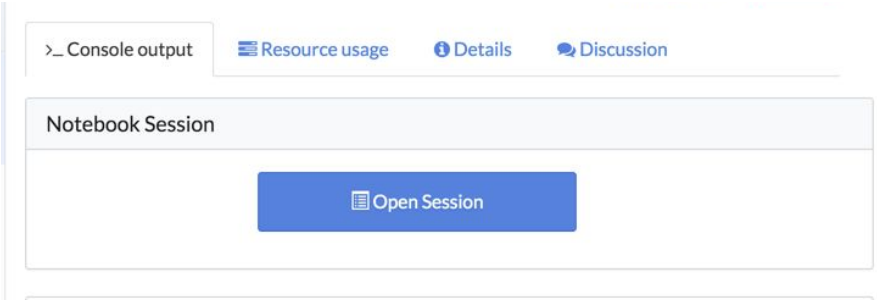
Click the `Stop` button in the screen's top right corner.



Important: Always remember to stop the run after your interactive sessions. Otherwise, that computer will continue running in the cluster and using up resources. For cloud and VPC users, runs that are not stopped also continue to incur charges.

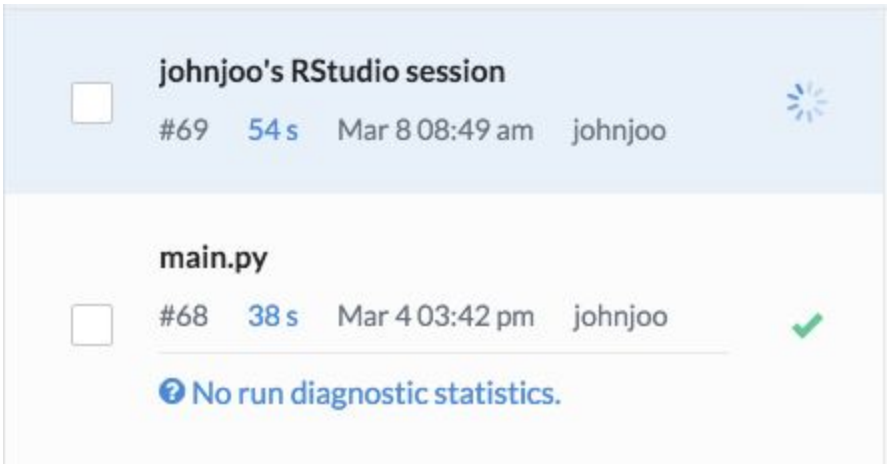
For R users

Select `RStudio` from the drop-down list. This will start a machine to host your RStudio session. In the right-hand panel, click `Open Session`.

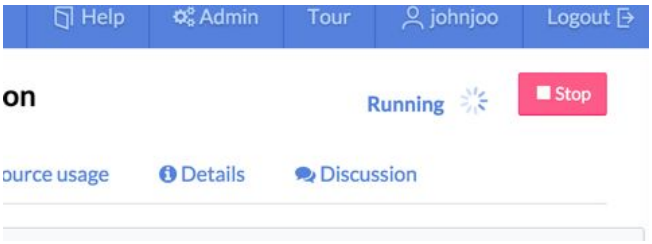


The RStudio session will open in a new browser. Open `analysis.R` to explore the script. Run the file and generate some initial results.

When you are done with the run, you will have to stop the session. Select the run for the RStudio session that is currently active.



Click the `Stop` button in the screen's top right corner.

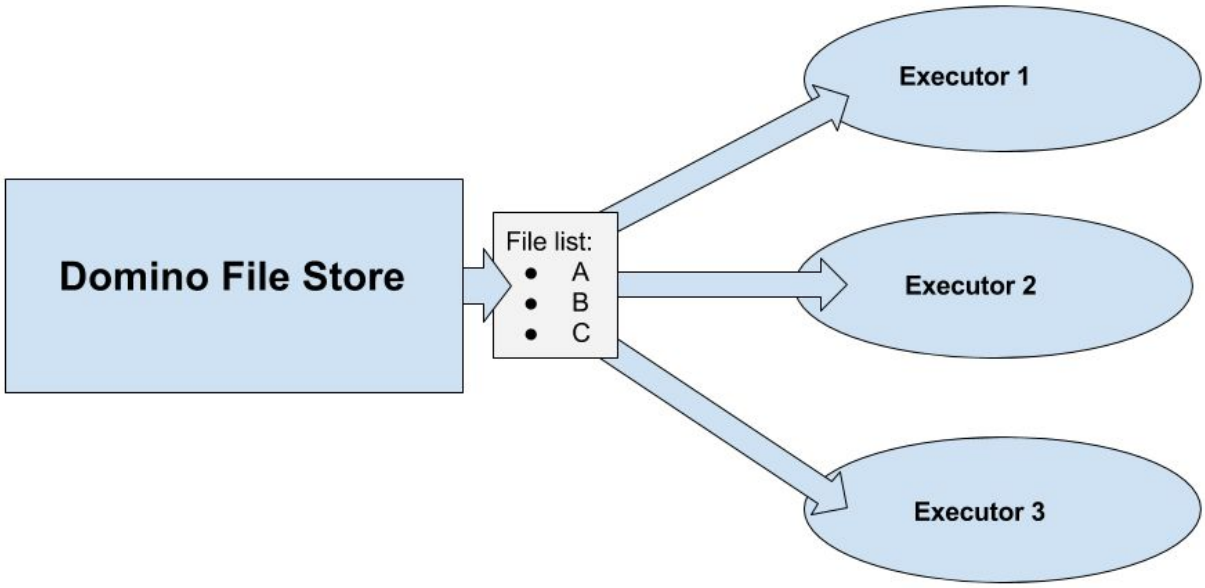


Always remember to stop the run after your interactive sessions. Otherwise, that computer will continue running in the cluster and using up resources.

A note about syncing in interactive sessions

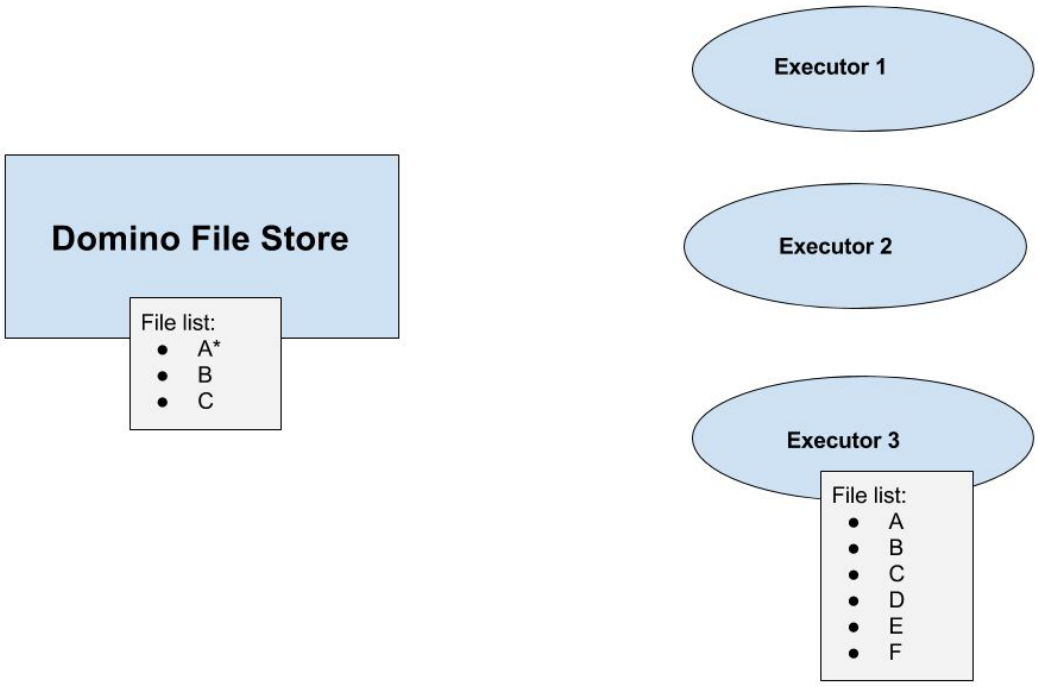
Once you have created and saved files in your interactive session, you will notice that those same files are not visible on the `Files` page. However, you do have access to them in your interactive session.

The files present on your `Files` page are hosted on the Domino Filestore, which is the source of all files and version history in your Domino project. All code execution happens on other machines called *executors*. When we started our new Jupyter session, we used an executor to host the interactive session. When you created new files in this Jupyter session, they were created on the executor.

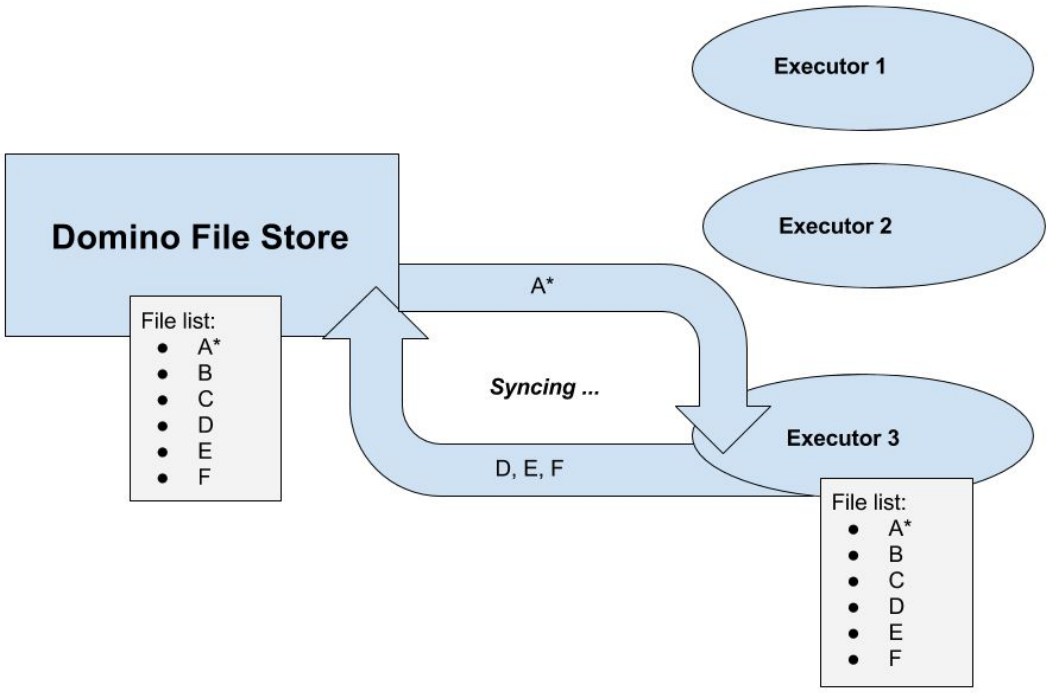


Domino automatically synchronizes the filestore and executors anytime the user stops an active session by clicking the `Stop` button. In the diagram above, the filestore pushes out files A, B and C to each of three executors.

During a run on Executor 3, three new files - files D, E and F - are generated. Meanwhile, file A has been changed on the filestore. The file lists on both machines are now depicted in the diagram below:

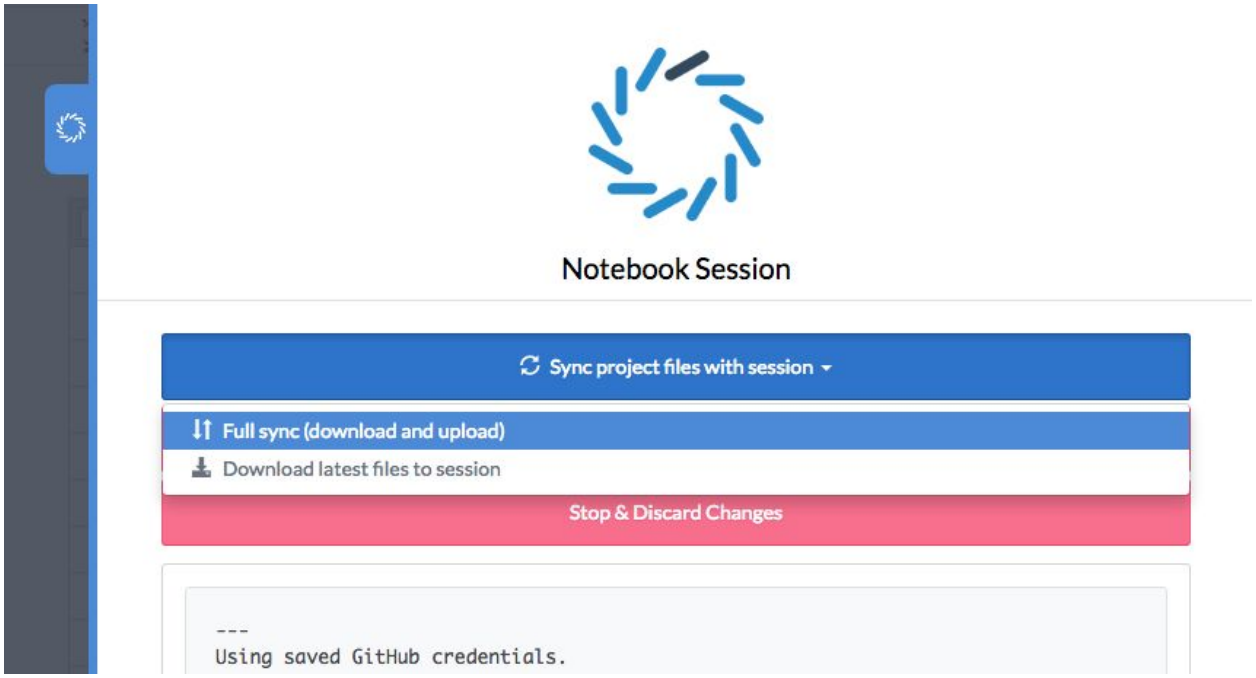


When Domino syncs the filestore and executor 3, it first uploads all new files from the executor to the filestore. It then downloads the files from the filestore to the executor. The process - and end result - are depicted in the diagram shown here:



It is also possible to manually sync these two machines. To do so, simply click the Domino pullover tab on the right-hand side of your screen.

Then click `Sync project files with session`, and select `Full sync (download and upload)`.



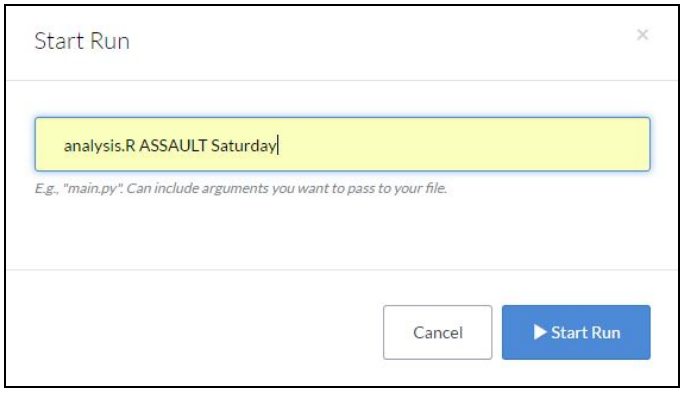
In this way, you can transfer files from the Domino Filestore to any Domino executor in an interactive session.

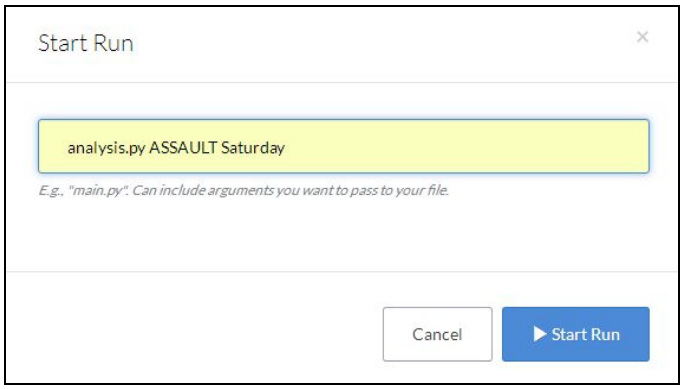
Running in Batch Mode

You can also start a run without first opening the file. To do so, first go back to the Domino browser tab. In the middle panel of the screen, click the `Run` button. A dialog box will appear.

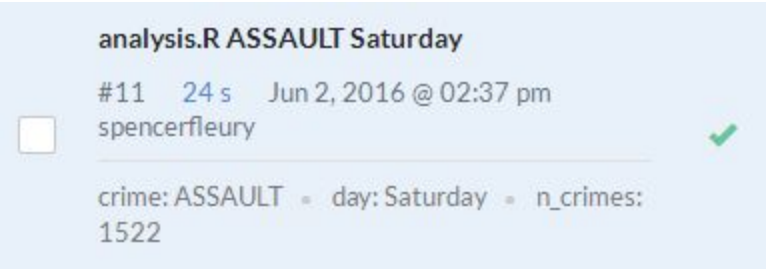
Type in the name of the file you want to run - `analysis.R` if you are working with R, `analysis.py` if you are working with Python - along with any additional parameters.

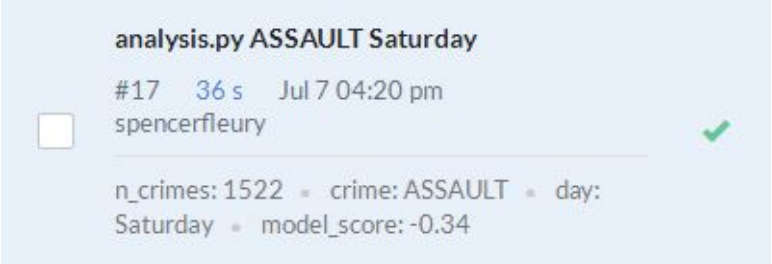
For this example, we will look at assaults that took place on Saturdays.



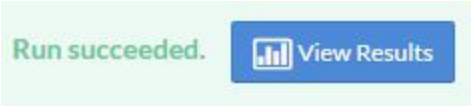


Click the `Start Run` button to launch the run. It should finish quickly.



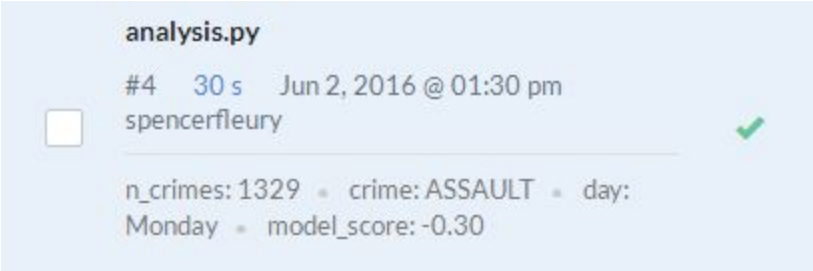


You can also view the results of the run directly, by clicking on the `View Results` button at the bottom of the right-hand panel of your screen.



Diagnostic statistics

At the conclusion of each run, Domino generates a set of diagnostic statistics.



By comparing these statistics generated from one run to the next, you can easily see whether your model development is making progress toward your goal, or whether your most recent changes have resulted in a less-focused model.

These diagnostic statistics are generated by code contained within the file used for the run.

```
In [ ]: #Run Diagnostic Statistics
        import json
        with open('dominostats.json', 'wb') as f:
            f.write(json.dumps({"model_score": '%.2f' %model.score(X), "n_crimes": len(X)}))
```

Python example of diagnostic statistics code

```
96. #Save some diagnostic statistics for later use
97. diagnostics = list("crime" = crime, "day" = day, "n_crimes" = nrow(filtered_crimedata))
98. library(jsonlite)
99. fileConn<-file("dominostats.json")
100. writeLines(toJSON(diagnostics), fileConn)
101. close(fileConn)
```

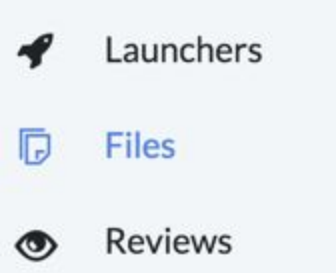
R example of diagnostic statistic code

This code can be modified and adjusted to provide a different set of statistics, or to display them in a different manner. For more information on these metrics, see http://support.dominodatalab.com/hc/en-us/articles/204348169-Run-Diagnostic-Statistics.

Editing files and retrieving previous versions of files

Domino tracks the changes you make to your files, and makes it easy to view those changes across versions.

In the left-hand sidebar, click `Files`. You will be returned to the `Files` tab.



Click on `analysis.py` (or `analysis.R`) to open it.

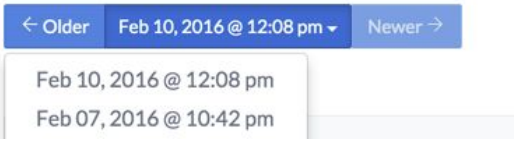
```
1. import folium
2. import pandas as pd
3. import sys
4. import matplotlib.pyplot as plt
5. import argparse
6.
7. SF_COORDINATES = (37.76, -122.45)
8. crimedata = pd.read_csv('SFPD_Incidents_-_Current_Year__2015_.csv')
9.
10. parser = argparse.ArgumentParser(description='Select crime and day.')
11. parser.add_argument('crime', nargs='?', default='ASSAULT')
12. parser.add_argument('day', nargs='?', default = 'Monday')
13.
14. args = parser.parse_args()
15. crimeselect = args.crime
16. dayselect = args.day
17.
18. def get_filtered_data(day, crime):
19.     dayselect = [day]
20.     crimeselect = crime
21.     daycond = crimedata['DayOfWeek'].isin(dayselect)
22.     crimecond = crimedata['Category'] == (crimeselect)
23.
24.     filtered_crimedata = crimedata[crimecond & daycond]
25.     return filtered_crimedata
26.
27. def get_n_crimes(day, crime):
```

Click on `Edit` to make changes to the file.

Now add a comment somewhere in the code.

```
32
33  crime_cat = crimedata.Category.unique()
34  print 'All categories of crimes:\n %s'%crime_cat
35
36  #Domino rocks!
37
38  #Exploration of where the crimes are of each category on specific day
39  #f,a = plt.subplots(7,6, sharex= True, sharey = True)
40  #f.set_size_inches(30, 30)
41
42  #for i, cat in enumerate(crime_cat):
```

Click `Save` to save the file. On the top of the screen, you will see a date. Click on the date.



These are all the versions of the `main.py` file. The comment you just added is in the most recent version. If you want to see the file as it previously existed - without the comment - just click on a previous version of the file.

```
31.
32. crime_cat = crimedata.Category.unique()
33. print 'All categories of crimes:\n %s'%crime_cat
34.
35.
36.
37. #Exploration of where the crimes are of each category on specific day
38. #f,a = plt.subplots(7,6, sharex= True, sharey = True)
39. #f.set_size_inches(30, 30)
40.
41. #for i, cat in enumerate(crime_cat):
```

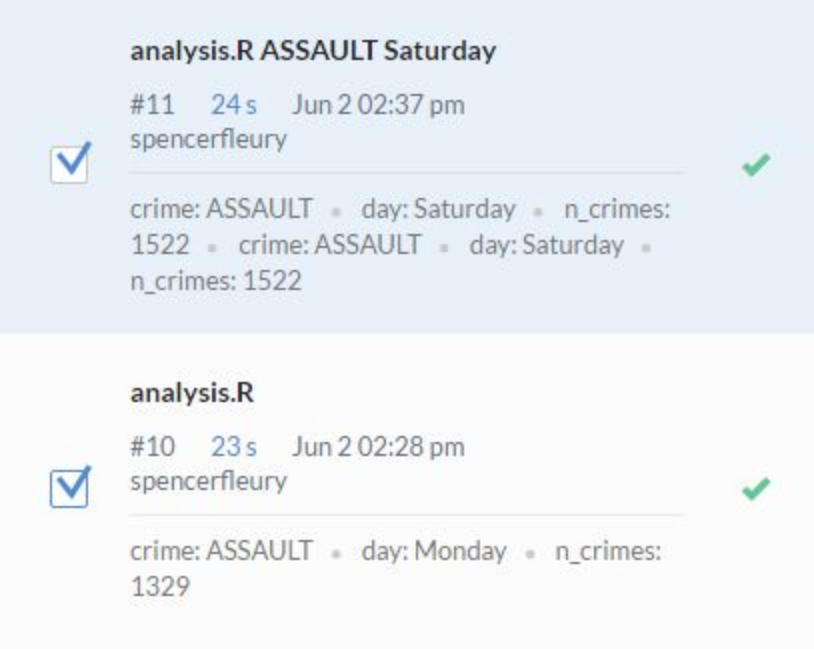
Notice the comment that was previously displayed in line 36 is now gone.

Compare two runs

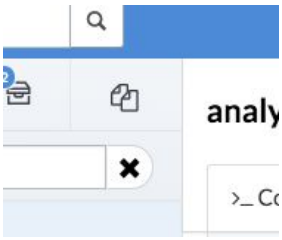
Once Domino has done multiple runs on the same dataset, it can be useful to see how those runs compare to each other.

To compare runs, click `Runs` in the left-hand sidebar. The `Runs` tab will open.

Click the checkboxes on the last two runs of `analysis.R`. If you only have one run, then repeat the process described above, substituting the word “Sunday” for “Saturday.”



After you’ve selected both runs, click on the icon that looks like two files overlapping.



This will open the `Comparing Runs` window.

From this view, you can scroll down to view the side-by-side differences in the output and input files.

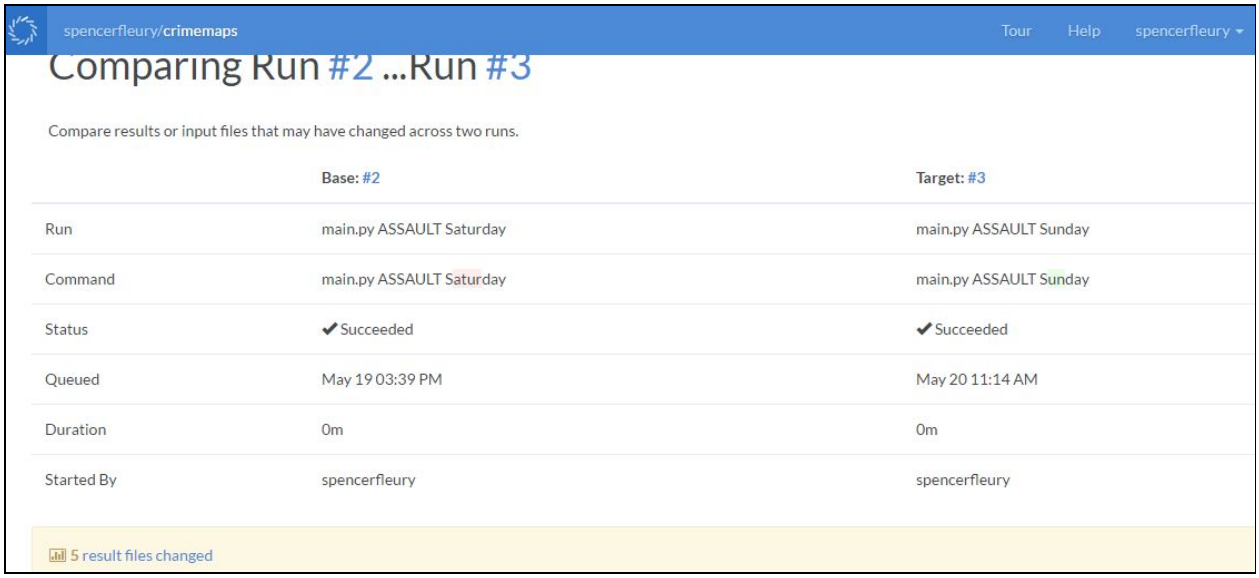
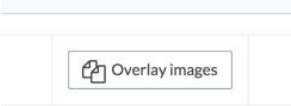

spencerfleury/crimemaps
Tour
Help
spencerfleury
Comparing Run #2 ...Run #3
Compare results or input files that may have changed across two runs.
Base: #2
Target: #3
Run
main.py ASSAULT Saturday
main.py ASSAULT Sunday
Command
main.py ASSAULT Saturday
main.py ASSAULT Sunday
Status
✔ Succeeded
✔ Succeeded
Queued
May 19 03:39 PM
May 20 11:14 AM
Duration
0m
0m
Started By
spencerfleury
spencerfleury
5 result files changed

Image Overlay

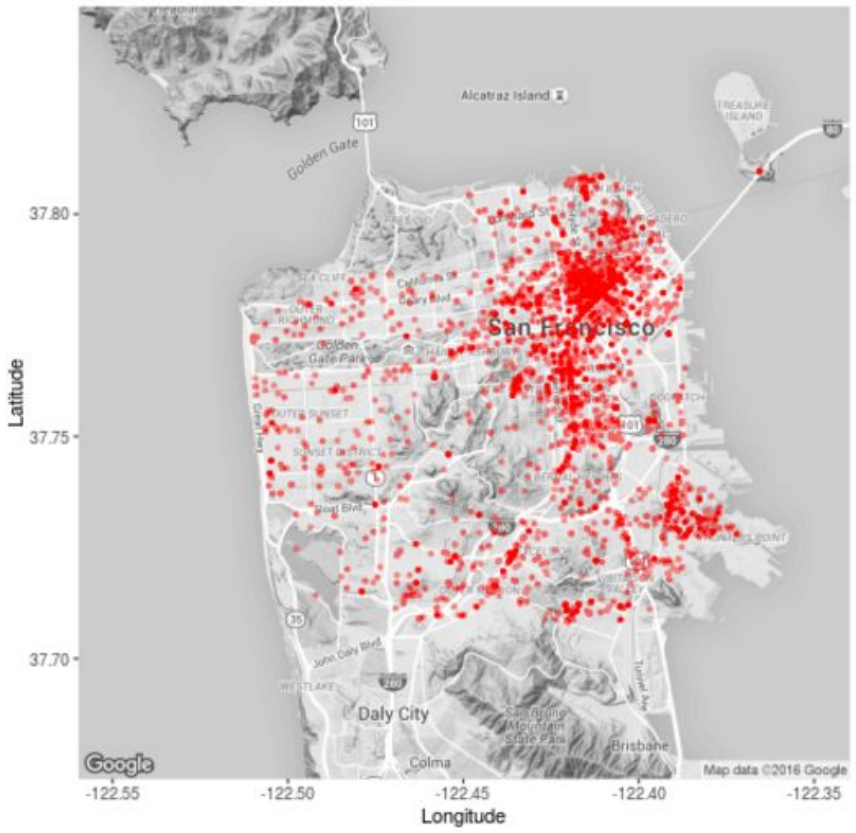
Throughout this guide, we have been working with data that have a geospatial component, and can be viewed in map form. With Domino, you can see the differences in run output presented in graphic form.

To generate this comparison, first click the `Overlay images` button.



To see the difference between these images, simply drag the scroll bar just below the map. The data shown will fade from the output of one run to the output of the other.

Image difference

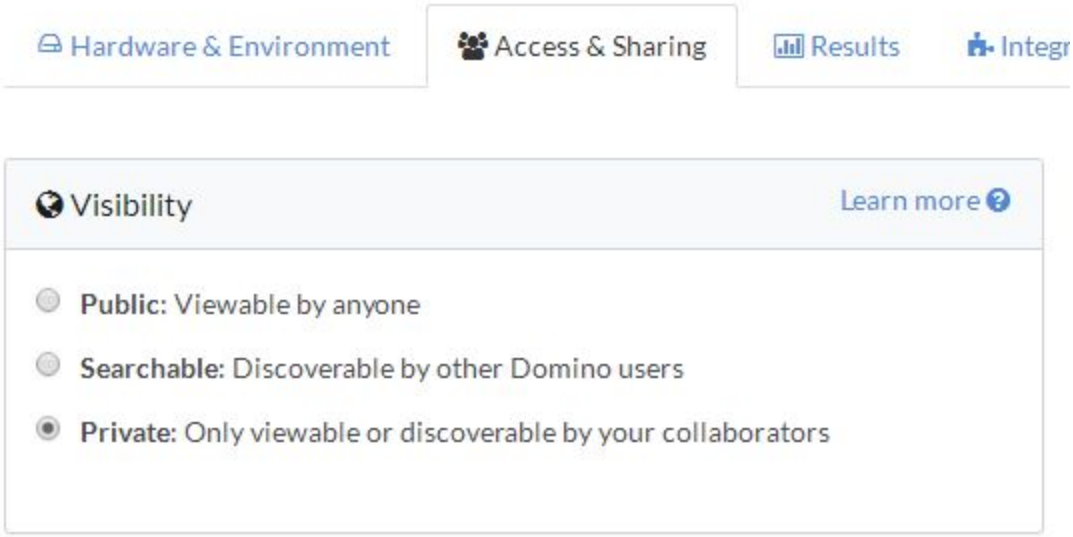


Base Target

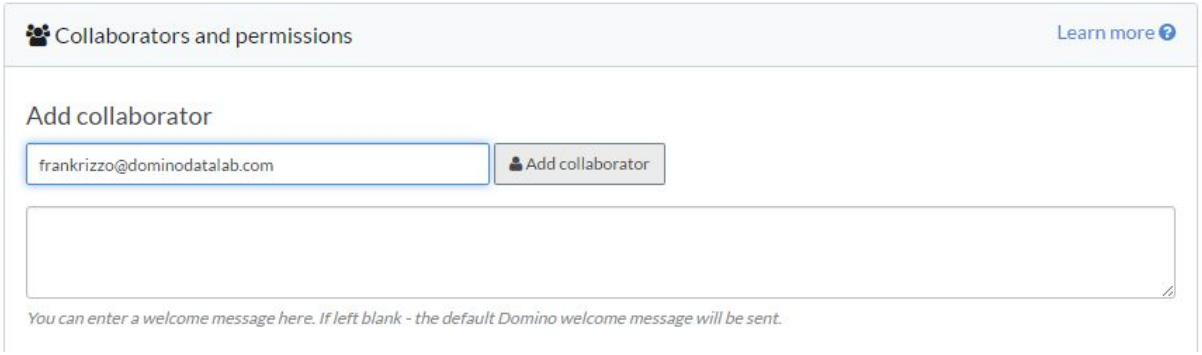
Invert colors on target (top) image

Setting your project's visibility

Domino allows you to choose who can see your project. Click the `Access & Sharing` tab to see your visibility options.



If you choose to set your visibility to `Private`, your project can only be found and viewed by your collaborators. To add collaborators to your project, go to the `Collaborators and permissions` section directly below the `Visibility` section.

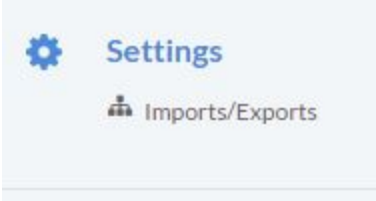


To add a collaborator, enter their username in the text box and click the `Add collaborator` button. They will receive an email inviting them to collaborate in your project.

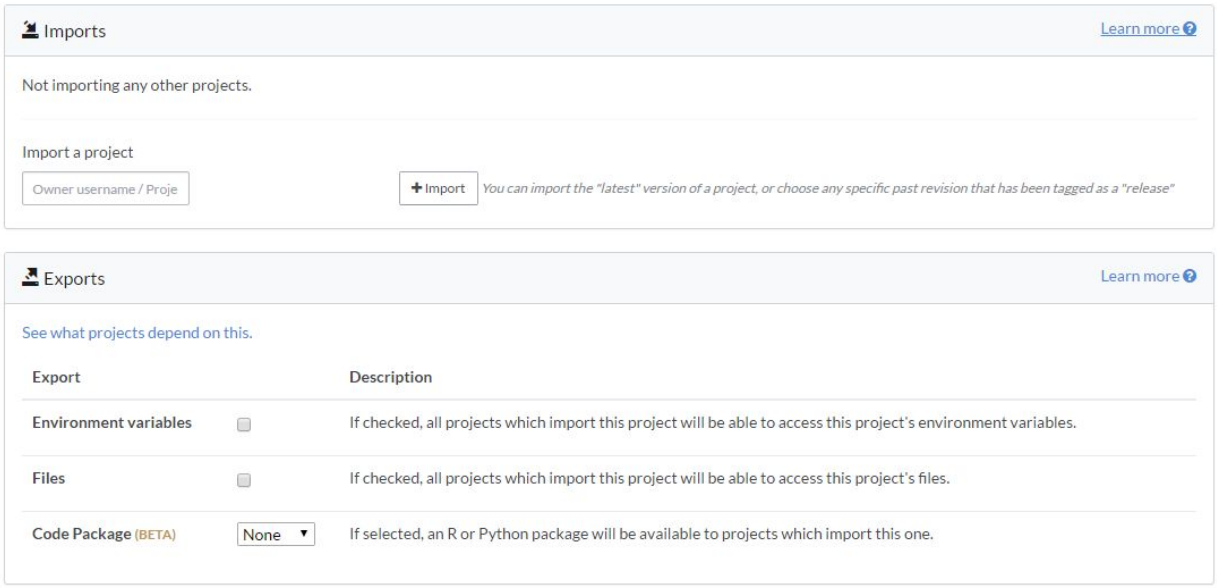
Imports and exports

Sometimes, you will want to re-use files or environmental variables from other projects in a new project. Domino allows you to set this up from the `Imports/Exports` screen.

In the left-hand sidebar, click `Imports/Exports` under `Settings`.



This will open the `Imports/Exports` window. Your screen should look like this:

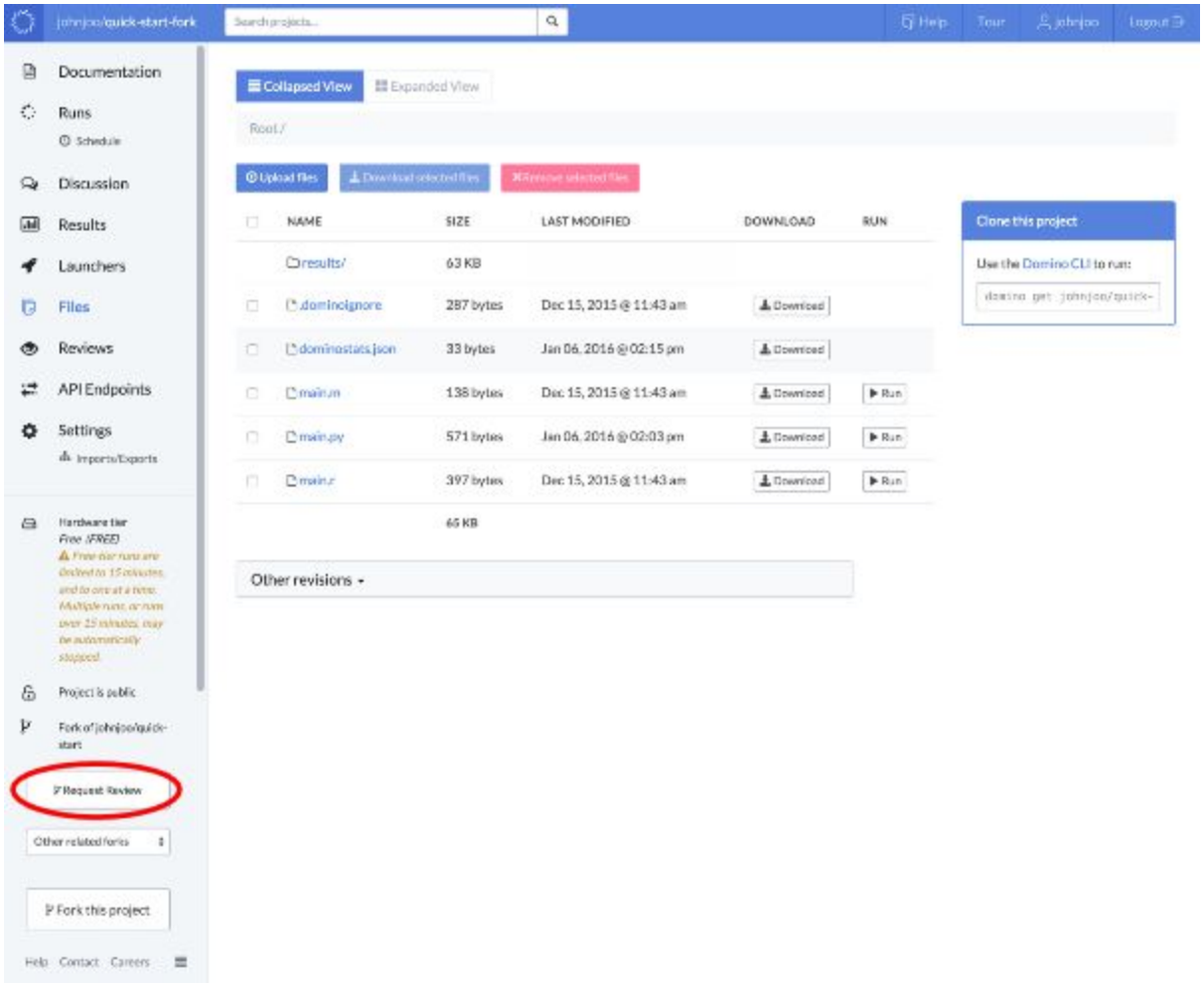


This is a feature you will be able to make use of once you've created or been granted access to more than one Domino project. For more information on how you can centralize and re-use configuration details, code, and data in Domino, see the official documentation: http://support.dominodatalab.com/hc/en-us/articles/205703615.

Reviews

Domino provides you with simple tools to manage forking and merging projects. The Reviews feature is an important component in the review process. You will have to request a review of any changes made--and that review will have to be accepted by the project owner--before any merges are completed.

To request a review prior to merging a project, click the `Request review` button in the left-hand sidebar.



For more details on the Reviews feature, see the official documentation: http://support.dominodatalab.com/hc/en-us/articles/207544483.

Command Line Interface (CLI)

By using the CLI, you will be able to link a folder on your local machine to a Domino project. Doing so will enable you to sync your files directly to Domino and start runs in Domino directly from the command line.

> **Protip:** We recommend giving identical names to your Domino project and to the folder that you are linking to Domino. This will prevent you from having to remember which Domino project your local folder is linked to.

Users who are more comfortable with a command line-based approach may prefer to use the CLI to interact with Domino quickly; additionally, some Domino features are currently only accessible through the CLI.

For more details on Domino's command line interface, see the official documentation: http://support.dominodatalab.com/hc/en-us/articles/204842905