



GA Data Science

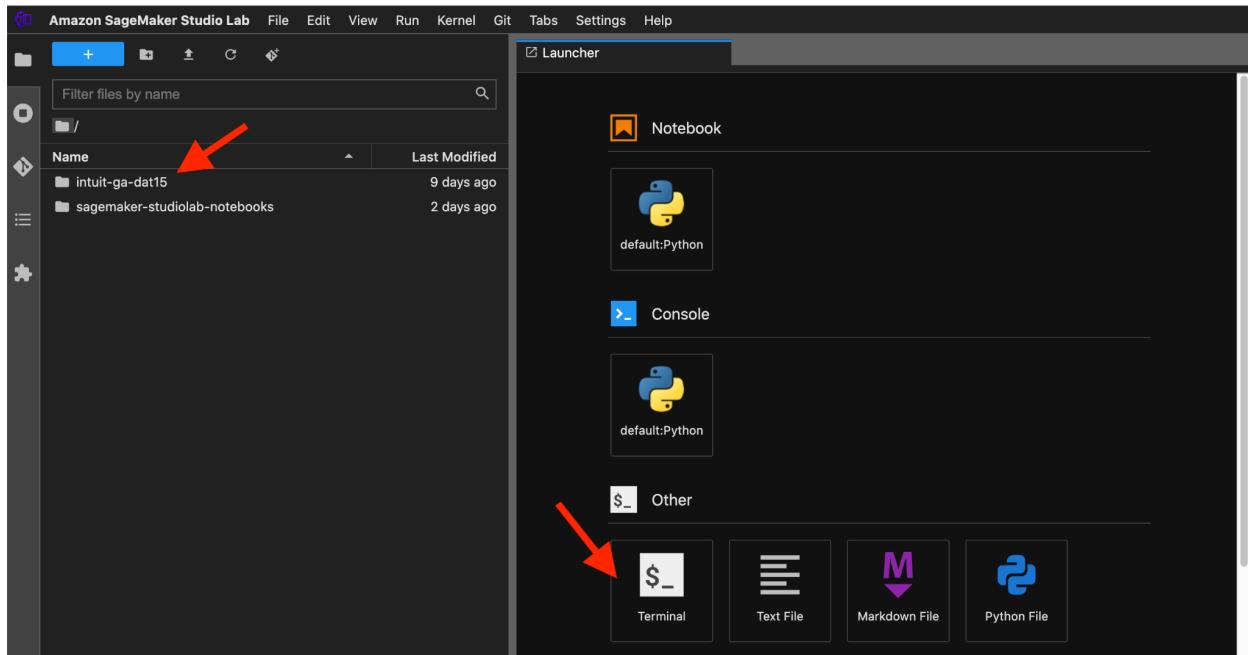
Intuit Cohort 15

Lesson 3: Project Hour

March 2, 2022

1. Clone the lecture repository to Sagemaker

Navigate to Sagemaker Studio. You should see the following. Remember that all of our work will be saved inside the folder “intuit-ga-dat15” that we created last time. If you wanted to, you could use the buttons on the far left side of the screen to do things like create new folders and clone git repositories. But that would be too easy (and a crutch). Instead, we’re going to do it the hard way, using terminal.



Open a new terminal session, and try the following commands.

```
cd intuit-ga-dat15
```

```
cd lectures
```

```
pwd
```

```
ls -l
```

You should see three repositories that we cloned the previous lessons: repos 00, 01, and 02. If the repos aren’t there, you probably missed this last class. But don’t worry! You can just use the same steps that we’re going to use for repo 03, and go back and get the previous repos as well.

```
(studiolab) studio-lab-user@default:~/intuit-ga-dat15/
(studiolab) studio-lab-user@default:~/intuit-ga-dat15$ cd lectures/
(studiolab) studio-lab-user@default:~/intuit-ga-dat15/lectures$ pwd
/home/studio-lab-user/intuit-ga-dat15/lectures
(studiolab) studio-lab-user@default:~/intuit-ga-dat15/lectures$ ls -l
total 0
drwxr-xr-x 5 studio-lab-user users 173 Feb 26 10:58 00-course-info
drwxr-xr-x 7 studio-lab-user users 190 Feb 26 10:58 01-welcome-dev-environment
drwxr-xr-x 5 studio-lab-user users 73 Feb 26 10:58 02-deploy-python-apps
(studiolab) studio-lab-user@default:~/intuit-ga-dat15/lectures$
```

Now navigate to our Data Science org in Github Enterprise.

<https://git.generalassemb.ly/intuit-ds-15>

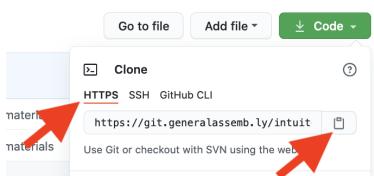
Find today's repository.

<https://git.generalassemb.ly/intuit-ds-15/03-exploratory-data-analysis>

Notice that the option to “Fork” has been disabled - this is by design. We don’t want you to fork this repo (because only the instructor should be modifying the content here).

Click on the green “Code” button.

A dialogue window will open. Choose “https” – do *not* choose “SSH” or “Github CLI”. Then copy the resulting URL using the clipboard icon to the right.



Now that you’ve copied the URL, go back to the Sagemaker console and your terminal there. Before we do the next steps, use the command **pwd** (print working directory) to make sure that you’re inside the right directory. When you clone a repo, it’s very important to keep track of the folder structure of your directories - you never want to accidentally clone a git repo inside of another repo! You should be here:

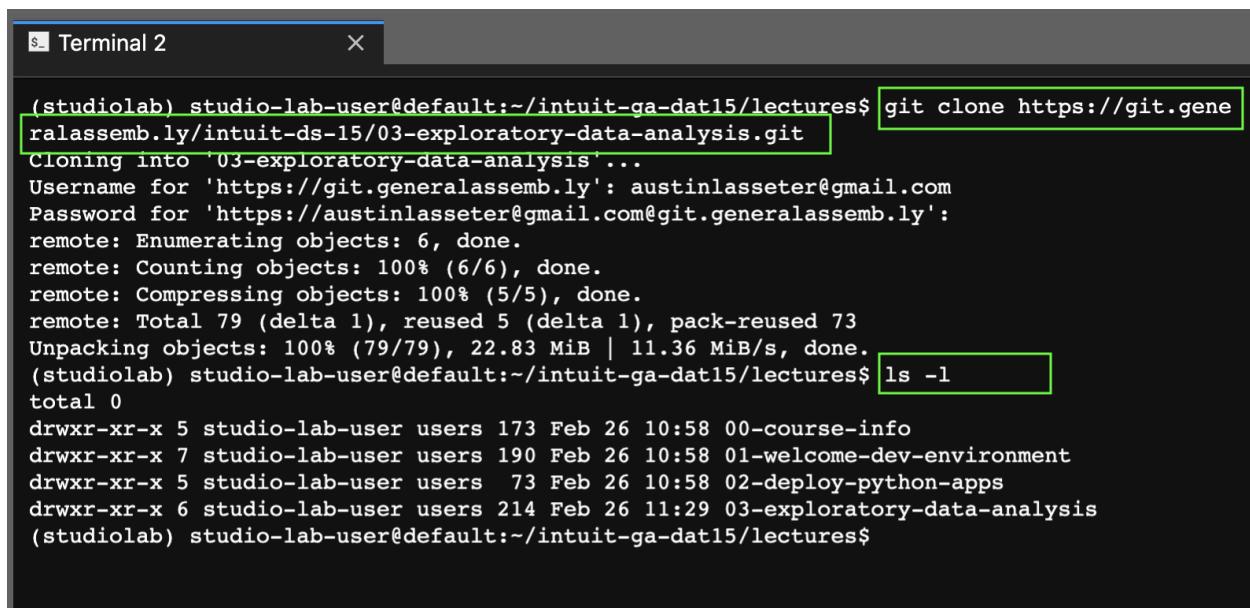
```
/home/studio-lab-user/intuit-ga-dat15/lectures
```

Use the following command to clone the repo to Sagemaker. After typing the words “git clone” you can just use ctrl+v to paste the URL that you copied in the previous step.

```
git clone https://git.generalassemblyly/intuit-ds-15/03-exploratory-data-analysis.git
```

Git will ask you for your username and your password. Note that github enterprise still uses a password, not a personal access token (but github public is the other way around). Also, when you try to type your password (or paste it using ctrl+V) the cursor in terminal doesn’t move at all - this is a security feature to protect your password, but it can be a little confusing the first time it happens. Just press enter after typing the password and all should be well.

You can then use the “list” command `ls -l` to show the files in your directory – the new 03 repo should now be present.



The screenshot shows a terminal window titled "Terminal 2". The user has run the command `git clone https://git.generalassemblyly/intuit-ds-15/03-exploratory-data-analysis.git`. The terminal then prompts for a username and password. After entering them, it shows the progress of cloning the repository, including object enumeration, counting, compressing, and unpacking. Finally, the user runs `ls -l` to list the contents of the cloned directory, which includes several files and folders corresponding to the repository's structure.

```
(studiolab) studio-lab-user@default:~/intuit-ga-dat15/lectures$ git clone https://git.generalassemblyly/intuit-ds-15/03-exploratory-data-analysis.git
Cloning into '03-exploratory-data-analysis'...
Username for 'https://git.generalassemblyly': austinlasseter@gmail.com
Password for 'https://austinlasseter@gmail.com@git.generalassemblyly':
remote: Enumerating objects: 6, done.
remote: Counting objects: 100% (6/6), done.
remote: Compressing objects: 100% (5/5), done.
remote: Total 79 (delta 1), reused 5 (delta 1), pack-reused 73
Unpacking objects: 100% (79/79), 22.83 MiB | 11.36 MiB/s, done.
(studiolab) studio-lab-user@default:~/intuit-ga-dat15/lectures$ ls -l
total 0
drwxr-xr-x 5 studiolab-user users 173 Feb 26 10:58 00-course-info
drwxr-xr-x 7 studiolab-user users 190 Feb 26 10:58 01-welcome-dev-environment
drwxr-xr-x 5 studiolab-user users 73 Feb 26 10:58 02-deploy-python-apps
drwxr-xr-x 6 studiolab-user users 214 Feb 26 11:29 03-exploratory-data-analysis
(studiolab) studio-lab-user@default:~/intuit-ga-dat15/lectures$
```

You can also navigate inside the 03 repo. Type “cd” for “change directory” and begin typing “03” then use the tab button to autocomplete. Check out the files there using the command “`ls -l`” and you’ll see that they are exact duplicates of the files listed in the remote origin.

As you can see in the screenshot below, the same files and folders are listed - but not in the same order. Terminal lists them alphabetically, while github lists folders at the top, files at the bottom. There’s also a hidden file (`.gitignore`) – you can show this in terminal with the command “`ls -a`” if you want. The `.gitignore` is sort of a trash bin that prevents garbage files on your machine (like `.pyc` files or `.DS_Store`) from ending up in github.

The screenshot shows a browser window with two tabs. The left tab is a JupyterLab terminal session titled '02-flying-cl' with the URL 'aws.studiolab/default/jupyter/lab'. The right tab is a GitHub repository page for 'intuit-ds-15/03-exploratory-data-analysis' with the URL 'https://git.generalassemb.ly/intuit-ds-15/03-exploratory-data-analysis'. The terminal session shows a command-line interface with several files listed in the current directory, including 'Exploratory_Data_Analysis.ipynb', 'Exploratory_Data_Analysis_Solution.ipynb', 'README.md', 'data', 'extra-materials', and 'slides'. The GitHub repository page shows a commit by 'austinlasseter' titled 'move slides' with a commit message: 'cleanup and rearrange materials'. The commit includes changes to 'data', 'extra-materials', and 'slides'. The repository has 1 branch, 0 tags, and 1 file named 'README.md'. A badge indicates the repository is 'active'.

Now your 03 repo is ready to go!

But we also want to go into the repos from last week – 00, 01, and 02 – and update the files there. This is because the instructor (yours truly) has made some changes to the remote origin and you want to update your clone with the latest version. We'll do this for one repo (02) and you can apply the same changes to the others.

To get out of the 03 directory, use the “cd” command (change directory) and two dots (which means go up one directory) like this: `cd ..`

Then navigate into the 02 repository using “cd” and begin typing “02” then use the tab button to autocomplete. Once inside, use the command “git pull” to obtain the latest version from the remote origin. It will once again ask for username and password.

The screenshot shows a browser window with a JupyterLab terminal session titled 'Terminal 2' with the URL 'aws.studiolab/default/jupyter/lab'. The terminal session shows a command-line interface with the user navigating to the '02-deploy-python-apps' directory and running the 'git pull' command. The command 'git pull' is highlighted with a green box. The terminal also shows other commands like 'pwd', 'ls -l', and 'cd ..'. The user is prompted for a 'Username for https://git.generalassemb.ly': [redacted].

Do the same for the other folders, 00 and 01. This same routine (git clone, git pull) should become a habit for you at the beginning of every lesson in our course.

2. Lecture materials

Once everyone is done cloning, we'll start the lecture. We'll cover the following materials, which are all included in the repo you just cloned.

- [Slide deck: AIML Terminology](#)
- [Slide deck: Exploratory Data Analysis With Pandas](#)
- [Python Notebook: Slide Notes](#)
- [Python Notebook: Exploratory Data Analysis Solutions](#)

This will take about 2 hours.

3. Clone the project repository to Sagemaker

Our course is structured so that we have lectures (which are saved in Github Enterprise) and projects (which are saved in Github Public). We do this because the lecture content is static (only the instructor will change it) but project content is dynamic (we want you to change and update the files as much as you want). This structure mirrors the usual way that most companies maintain their codebase for private and public use.

Some of the steps will be the same here, but some will be different. Get into the habit of following these same steps at the beginning of every lesson.

In Sagemaker, navigate up to the project folder and see what's there. If you successfully cloned the repos last week, you should see folders 101 and 102. If not, that's okay! You can just follow the steps we're about to use for the new repo.

```
(studiolab) studio-lab-user@default:~/intuit-ga-dat15/lectures/02-deploy-python-apps$ pwd
/home/studio-lab-user/intuit-ga-dat15/lectures/02-deploy-python-apps
(studiolab) studio-lab-user@default:~/intuit-ga-dat15/lectures/02-deploy-python-apps$ cd ../../
(studiolab) studio-lab-user@default:~/intuit-ga-dat15/projects$ ls -l
total 0
drwxr-xr-x 4 studio-lab-user users 111 Feb 16 18:51 101-static-website-example
drwxr-xr-x 5 studio-lab-user users 181 Feb 24 01:50 102-flying-dog-beers
(studiolab) studio-lab-user@default:~/intuit-ga-dat15/projects$
```

Now navigate to our projects org in Github Public. <https://github.com/plotly-dash-apps>

This is a good time to point out that throughout this course, we'll be using an open source python library called "plotly dash" – it's a popular tool for building data science applications, and a replacement for the older python app library known as "flask". Read all about it here:

<https://dash.plotly.com/introduction>

I've tried to include a lot of repos inside this org - these are designed to be simple introductory apps for students to play with and learn. The basic structure of the org includes simple and intermediate examples (nothing in this org should be considered "advanced").

- 100 series: simple, static dash applications
- 200 series: simple interactive apps that use a "callback" function
- 300 series: simple apps that incorporate the "pandas" library for interacting with datasets
- 400 series: intermediate callback functions without datasets
- 500 series: intermediate apps with datasets and simple machine learning
- 600 series: apps with intermediate machine learning

For today's project, we're going to focus on the following repo:

<https://github.com/plotly-dash-apps/203-radio-callbacks>

This is what we're going to build in today's project. You can view the finished app here:

<https://dash-radio-callback.herokuapp.com/>

For now, let's focus on cloning the repo. The first thing you should do is check out your github settings and make sure you have a personal access token - we created this in last week's lesson. If you need a review, you can see those instructions here:

<https://git.generalassemb.ly/intuit-ds-15/01-welcome-dev-environment/blob/master/Setting%20up%20AWS%20Sagemaker%20Studio%20Lab.pdf>

You can verify that you have a personal access token under github settings/developer settings, but you can't actually see the token itself - you had to copy it to a secure location when you first created it. If you missed that last week, no worries - you can just create a new token now, using the "generate new token" option on the right-hand side.

The screenshot shows the GitHub developer settings page at github.com/settings/tokens. The 'Personal access tokens' tab is selected. It displays a table with one row for a token named 'github projects — repo'. The token was last used 'Within the last 2 weeks' and expires on 'Sun, Feb 12 2023'. Buttons for 'Generate new token' and 'Revoke all' are visible.

Okay, back to our repo <https://github.com/plotly-dash-apps/203-radio-callbacks>. Now let's "fork" the repo. We didn't fork the lecture repo in Github Enterprise because you will never modify the lecture materials - only the instructor does that. But you can and should modify the contents of the project folder! Forking is github's way of allowing you to take ownership of the project materials.

First, notice that the name of the repo includes the org "plotly-dash-apps" as a prefix. That's about to change. Look for the "fork" button in the top right corner, and click it.

The screenshot shows a GitHub repository page for 'plotly-dash-apps/203-radio-callbacks'. The 'Fork 0' button in the top right is highlighted with a green box. The repository details show 1 branch, 0 tags, and 3 commits from 'austinlasseter'. The 'About' section notes 'No description, website, or topics provided.'

It will ask you where you want to fork it. Choose your personal account.

The screenshot shows a modal window titled 'Fork 203-radio-callbacks'. It asks 'Where should we fork 203-radio-callbacks?' and lists 'austinlasseter' as an option.

You should now see a new version of the repo, where the prefix is your personal github username, like this:

This screenshot shows a GitHub repository page for 'austinlasseter / 203-radio-callbacks'. The repository is public and was forked from 'plotly-dash-apps/203-radio-callbacks'. The main branch is 'main'. A message in the center of the page states: 'This branch is up to date with plotly-dash-apps:main.'

There's also a message letting you know that your fork is up to date with the "main" branch of the original repo. As time goes on, your fork will begin to diverge from the origin more and more. That's okay. Github has lots of ways to help you keep things more or less in sync ([like this](#)), but that's outside the scope of what we want to do today.

Now let's clone to Sagemaker. Click on the green "code" button and choose "https" and then the "copy to clipboard" icon.

A modal window titled 'Clone' is displayed, showing the URL 'https://github.com/austinlasseter/203-' and a copy icon. A green arrow points to the 'Code' button in the top right corner of the GitHub interface.

Back in Sagemaker terminal, make sure that you're in the appropriate directory by using the "pwd" command. You should be here:

```
/home/studio-lab-user/intuit-ga-dat15/projects/
```

Clone the repo. Use **ctrl+v** to paste the url (obviously it will have your name, not mine).

```
git clone https://github.com/austinlasseter/203-radio-callbacks.git
```

Because this is set to be a public repo, git will not ask you for your username or personal access token when you clone the repo - but it will ask you for this information when you try to

commit your changes using the “git push” command. More on that in the next section. For now, just “cd” into the repo and check out the files there. They will match what’s in the remote origin.

The screenshot shows two side-by-side windows. On the left is a terminal window titled 'Terminal 2' showing the command 'git clone https://github.com/plotly-dash-apps/203-radio-callbacks'. The terminal output shows the cloning process, including object enumeration and unpacking. On the right is a GitHub repository page for 'austinlasseter / 203-radio-callbacks'. A green box highlights the file listing on the right, which includes 'assets', '.gitignore', 'LICENSE', 'Procfile', 'README.md', 'app.py', 'requirements.txt', and 'runtime.txt'. A green arrow points from the terminal's file listing to the GitHub file listing, indicating they are identical.

That's it for now! You've successfully cloned the project repo.

4. Setting up your virtual environment with “conda”

When you're working on an application and you plan to deploy it to the cloud, you need to get in the habit of always working in a python environment. This can be a little tedious at first, but it helps make sure that the list of libraries that are installed on one machine will be identical to the list of libraries installed on another. Failing to do this will result in a lot of conflict down the road.

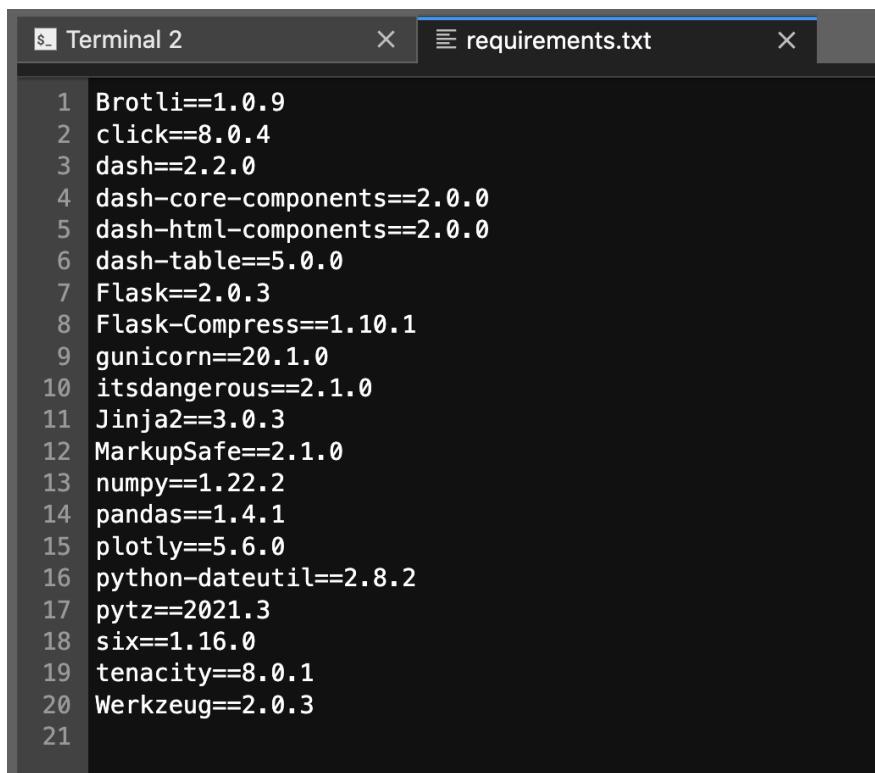
There are two ways to main virtual environments - pip and conda. Sagemaker is built using the Anaconda distribution of python (so it uses conda), but it's also compatible with pip environments. However, the heroku platform is not compatible with Anaconda environments, and your requirements.txt file will choke on heroku if you create it using conda. So let's use pip.

Because we're going to be working a lot with Heroku, we also need to be aware of which python versions Heroku supports (it only supports a few versions). You can read more about that here: <https://devcenter.heroku.com/articles/python-support>

For the purpose of today's lesson, here's what you should do.

#	Instructions	Terminal Command
1	Confirm that Sagemaker already has installed python version 3.9.7	python --version
2	Create a new environment named simple-dash-env	python -m venv env
4	Activate the new environment to use it	source env/bin/activate
5	Install the “dash” library	pip install dash
6	Install the “pandas” library	pip install dash
7	Install the “gunicorn” library	pip install gunicorn
8	Create the requirements.txt file	pip freeze > requirements.txt

Take a moment to check out the new requirements file that's been created. It lists all of your python dependencies which are installed in your environment. It will also be used to install these dependencies on heroku. This is a pretty basic list - for more advanced apps, you'll need to have more libraries installed.



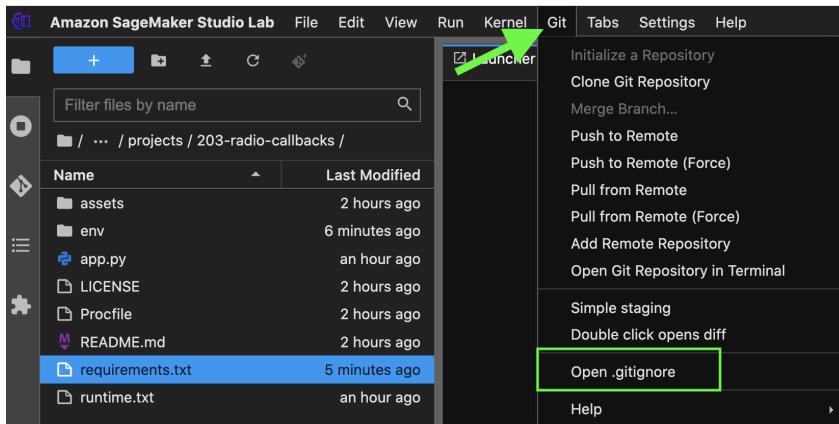
```

1 Brotli==1.0.9
2 click==8.0.4
3 dash==2.2.0
4 dash-core-components==2.0.0
5 dash-html-components==2.0.0
6 dash-table==5.0.0
7 Flask==2.0.3
8 Flask-Compress==1.10.1
9 gunicorn==20.1.0
10 itsdangerous==2.1.0
11 Jinja2==3.0.3
12 MarkupSafe==2.1.0
13 numpy==1.22.2
14 pandas==1.4.1
15 plotly==5.6.0
16 python-dateutil==2.8.2
17 pytz==2021.3
18 six==1.16.0
19 tenacity==8.0.1
20 Werkzeug==2.0.3
21

```

We've also got a new folder called “env” which contains all the binary files used to install these libraries - that's going to slow things down a lot when we try to push to github. To avoid that logjam, we need a hidden file called “.gitignore” which will tell git to “ignore” everything in the

/env/ folder. But for reasons no one is sure about, JupyterLab won't display hidden files in the left-hand navigation. Instead you have to open it using the "git" tab in the ribbon.



Open up the .gitignore and make sure that /env/ is listed at the top as "env/" – I've provided a template with a lot of other common garbage files as well.

A screenshot of a code editor window titled 'Launcher' with the file '.gitignore' open. The code editor has a dark theme. The file contains the following content:

```
1 env/
2
3 # Byte-compiled / optimized / DLL files
4 __pycache__/
5 *.[py|c|o]
6 *$py.class
7
8 # C extensions
9 *.so
10
11 # Distribution / packaging
12 .Python
13 build/
14 develop-eggs/
15 dist/
16 downloads/
17 eggs/
18 .eggs/
19 lib/
```

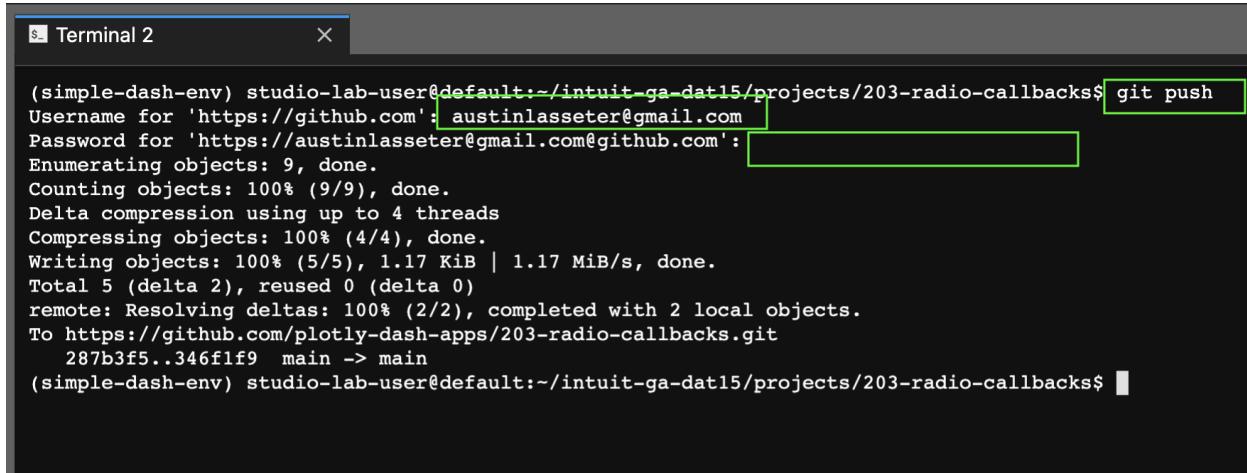
The first line, 'env/', is highlighted with a green box. The rest of the file follows the standard .gitignore template for Python distribution files.

Okay, that was easy. Let's move on.

9	Test out the file "app.py"	python app.py
10	Check to see that the following message is displayed in terminal:	Dash is running on http://127.0.0.1:8050/
11	Stop running the file "app.py"	keyboard: ctrl+c
12	Add your changes to git	git add .
13	Write a commit message	git commit -m "I updated the

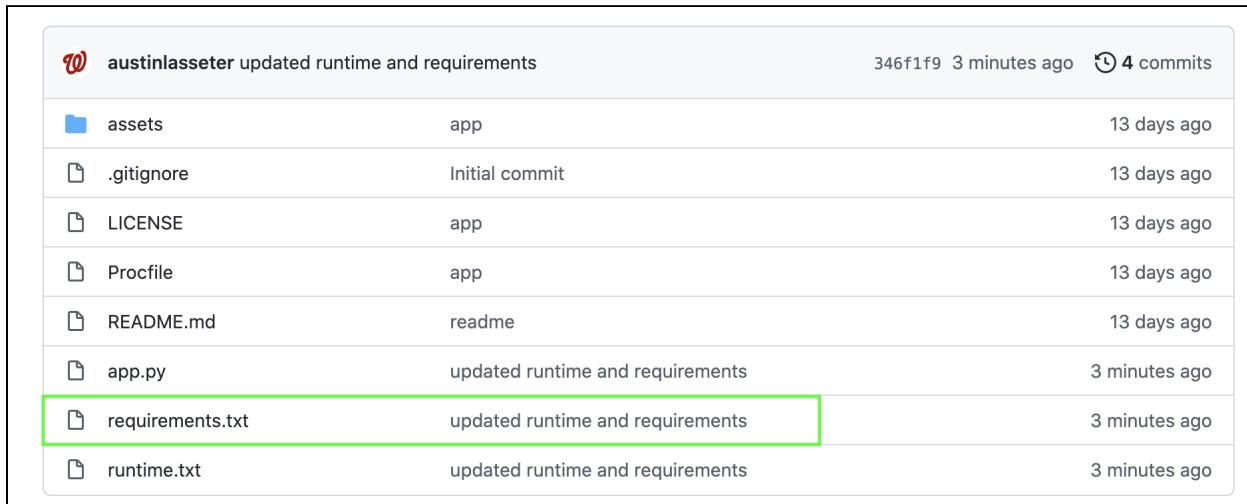
		requirements file"
14	Push your changes to the remote origin	git push

At this point, github will ask you for your username and your “password” (which is actually your personal access token). Remember that the cursor won’t move while you type (or paste) the token, so just press “enter” when you’re done. It should look like this:



```
(simple-dash-env) studio-lab-user@default:~/intuit-ga-dat15/projects/203-radio-callbacks$ git push
Username for 'https://github.com': austinlasseter@gmail.com
Password for 'https://austinlasseter@gmail.com@github.com':
Enumerating objects: 9, done.
Counting objects: 100% (9/9), done.
Delta compression using up to 4 threads
Compressing objects: 100% (4/4), done.
Writing objects: 100% (5/5), 1.17 KiB | 1.17 MiB/s, done.
Total 5 (delta 2), reused 0 (delta 0)
remote: Resolving deltas: 100% (2/2), completed with 2 local objects.
To https://github.com/plotly-dash-apps/203-radio-callbacks.git
  287b3f5..346f1f9 main -> main
(simple-dash-env) studio-lab-user@default:~/intuit-ga-dat15/projects/203-radio-callbacks$
```

Back in your remote origin repo on github, you should see your commit message similar to this:

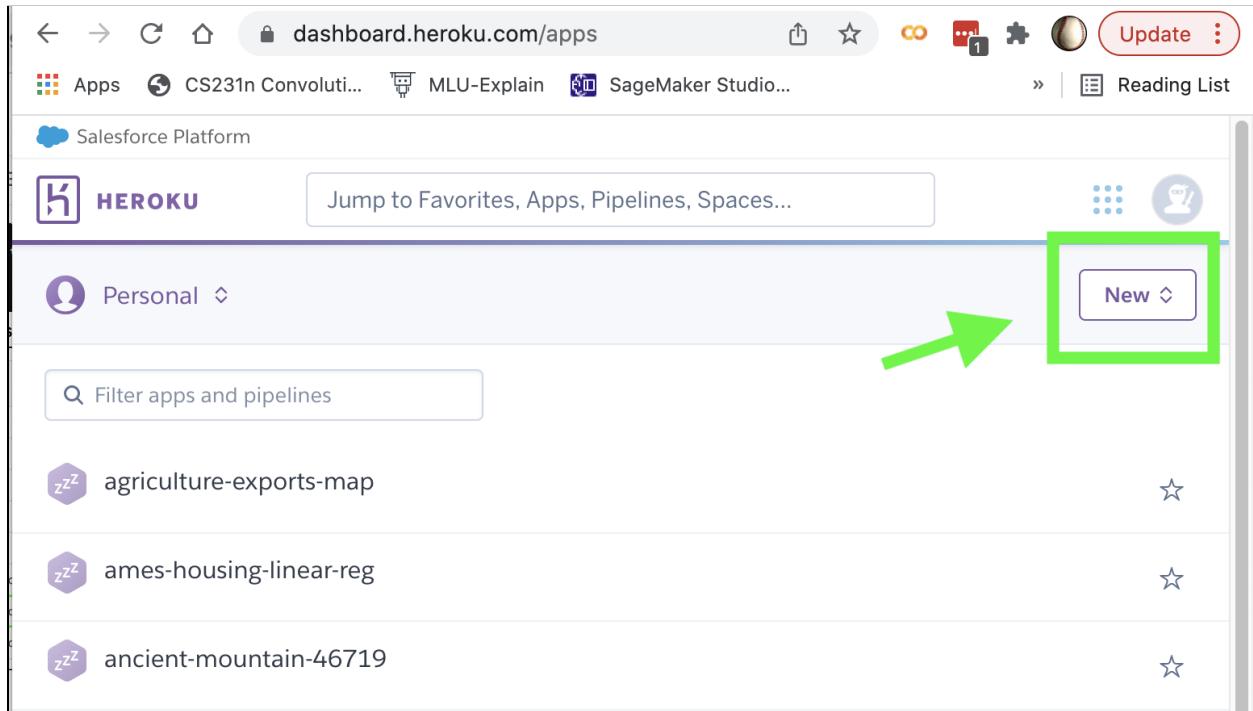


File	Description	Time
austinlasseter updated runtime and requirements	346f1f9 3 minutes ago	4 commits
assets	app	13 days ago
.gitignore	Initial commit	13 days ago
LICENSE	app	13 days ago
Procfile	app	13 days ago
README.md	readme	13 days ago
app.py	updated runtime and requirements	3 minutes ago
requirements.txt	updated runtime and requirements	3 minutes ago
runtime.txt	updated runtime and requirements	3 minutes ago

Keep in mind that you only have to update the requirements file when you install a new python library into your environment. It’s a good idea to use a fresh environment whenever you start a new project, because typically a new project means new dependencies. But you’ll get a better idea of that as you try out more and more projects using Plotly Dash.

5. Deploying your app to Heroku

Now let's try deploying our application on heroku. Go to the heroku website and start a new app.



Keep in mind that we're using the free tier so there are some limits. If you are “unverified” (i.e., you did not give them your credit card) then you are limited to 5 apps. If you verify the account by adding a credit card on file, you can have up to 100 apps on the (verified) free plan, and still the 5 app limit for unverified. There is no charge as long as you maintain your apps on the free-tier dynos. You can read more about this here:

- <https://stackoverflow.com/questions/48524995/how-many-apps-are-allowed-on-a-hobby-heroku-plan>
- <https://www.heroku.com/free>
- <https://devcenter.heroku.com/articles/limits>

Give your app a unique name. This one is mine.

Create New App

App name

xkcd-radio-callback
 ✓

xkcd-radio-callback is available

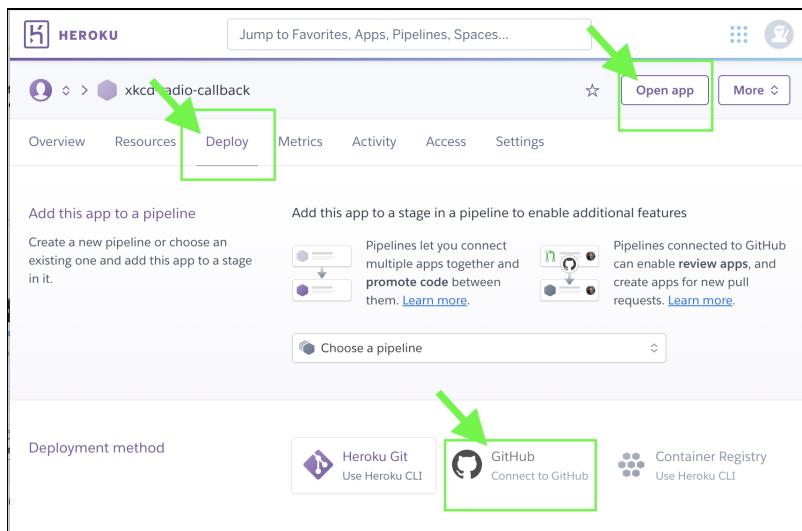
Choose a region

United States
▼

[Add to pipeline...](#)

Create app

You'll see a lot of options in the next window. Feel free to explore a little. If you choose "open app" you'll see that a placeholder website has been created for you, but nothing is there yet. Under "deploy" choose "github". (There are other ways to deploy to github, like the CLI, but Sagemaker is not compatible with them).



The screenshot shows the Heroku dashboard for the 'xkcd-radio-callback' app. At the top, there's a navigation bar with the Heroku logo, a search bar, and various icons. Below it, the app name 'xkcd-radio-callback' is displayed with a dropdown arrow. The dashboard features several tabs: Overview, Resources, Deploy (which is highlighted with a green box and arrow), Metrics, Activity, Access, and Settings. To the right of the tabs are 'Open app' and 'More' buttons. The main content area includes sections for adding the app to a pipeline and deploying via GitHub. The 'Deployment method' section at the bottom shows three options: Heroku Git, GitHub (which is highlighted with a green box and arrow), and Container Registry.

Search for your repo, and choose “connect”. The first time you do this, heroku will ask you for your username and personal access token.

The screenshot shows the Heroku Dashboard interface. On the left, there's a section titled "Connect to GitHub" with the sub-instruction: "Connect this app to GitHub to enable code diffs and deploys." To the right, there's a search bar labeled "Search for a repository to connect to" with the placeholder "austinlasseter". Below the search bar, it says "203-radio callbacks" and has a "Search" button. A note below the search bar says "Missing a GitHub organization? [Ensure Heroku Dashboard has team access](#)." At the bottom right of this section is a "Connect" button next to the repository name "austinlasseter/203-radio callbacks".

If you authenticate successfully, you will see the following message.

The screenshot shows the Heroku Dashboard after a successful connection. On the left, it says "App connected to GitHub" and "Code diffs, manual and auto deploys are available for this app." To the right, it displays a summary: "Connected to [austinlasseter/203-radio callbacks](#) by [austinlasseter](#)". It includes a "Disconnect..." button. Below this, there's a note: "Releases in the [activity feed](#) link to GitHub to view commit diffs".

Now scroll down to “manual deploy” and click “deploy branch”.

The screenshot shows the "Manual deploy" section of the Heroku Dashboard. It starts with the heading "Manual deploy" and the instruction "Deploy the current state of a branch to this app." To the right, it says "Deploy a GitHub branch" and "This will deploy the current state of the branch you specify below. [Learn more](#)". Below this, it says "Choose a branch to deploy" and shows a dropdown menu set to "main". To the right of the dropdown is a "Deploy Branch" button.

Wait a couple of minutes for heroku to do its thing. When your app is ready, you should see the option to open your application. My version is deployed here:

<https://xkcd-radio-callback.herokuapp.com/>

Troubleshooting deployment: Sometimes, things do go wrong. You can view your error messages in the log files under “Activity”:

The screenshot shows the Heroku dashboard for the application 'xkcd-radio-callback'. The top navigation bar includes 'Personal', 'GitHub' (linked to 'austinlasseter/203-radio-callback'), and tabs for 'Overview', 'Resources', 'Deploy', 'Metrics', 'Activity' (which is highlighted with a green box and has a green arrow pointing to it), 'Access' (with a green arrow pointing to the 'View build log' link), and 'Settings'. The 'Activity Feed' section lists four events:

- austinlasseter@gmail.com: Deployed 06fd9227 Today at 9:31 AM · v3
- austinlasseter@gmail.com: Build succeeded Today at 9:30 AM · [View build log](#)
- austinlasseter@gmail.com: Build failed Today at 9:06 AM · [View build log](#)
- austinlasseter@gmail.com: Build failed Today at 8:56 AM · [View build log](#)

There are usually common problems when you deploy to Heroku:

- There's a syntax error in your python code (and the logs will tell you the line number where you need to make that fix)
- There's a missing dependency in your "requirements.txt" file (so you need to rebuild your virtual environment, and freeze your requirements).

Of course lots of other problems can and will pop up, and we'll try to handle those on a case-by-case basis. Here are some links to help you sort things out:

- Plotly's [Dash deployment guide](#)
- Heroku's [deployment guide](#)
- Fantastic [blog post](#) that dives deep
- Excellent [YouTube tutorial](#)

6. Make some changes and deploy again

Okay, back in Sagemaker let's make some simple changes to the file "app.py". Because we want to keep things simple today, we are only going to change lines 10-15.

```

7
8 ##### Define your variables #####
9
10 myheading1='How to use callbacks'
11 tabtitle = 'xkcd'
12 list_of_options=['box plot', 'correlation', 'git commit', 'scatterplot']
13 list_of_images=['outlier.png', 'correlation.png', 'gitcommit.jpg', 'scatterplot.png', 'good_code.png']
14 sourceurl = 'https://xkcd.com/'
15 githublink = 'https://github.com/plotly-dash-apps/203-radio-callbacks'
16

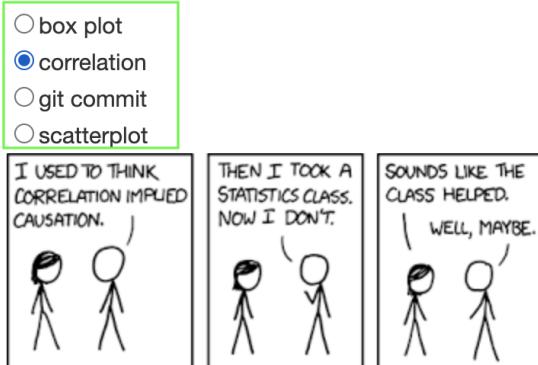
```

It should be simple to update the variables - just make sure that you keep all your strings in quote marks.

- myheading1
- tabtitle
- sourceurl (replace this with your own data source)
- githublink (replace this with your personal github repo)

The “list of options” is what will appear in your radio-button options. This will get used further down in lines 33-36. Notice that there are exactly four options in line 12, and four images in line 13. Let’s keep the list at exactly four items (no more, no less) for right now.

How to use callbacks

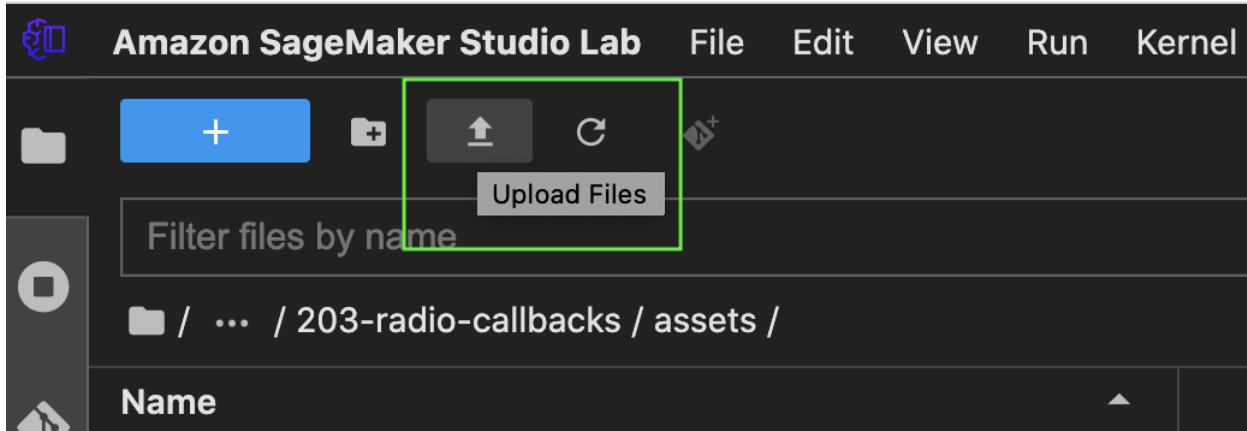


The images are all stored in another folder, called “assets”. Let’s take a look at its contents.

Name	Last Modified
chart.jpg	2 hours ago
correlation.png	2 hours ago
favicon.ico	2 hours ago
gitcommit.jpg	2 hours ago
good_code.png	2 hours ago
outlier.png	2 hours ago
scatterplot.png	2 hours ago

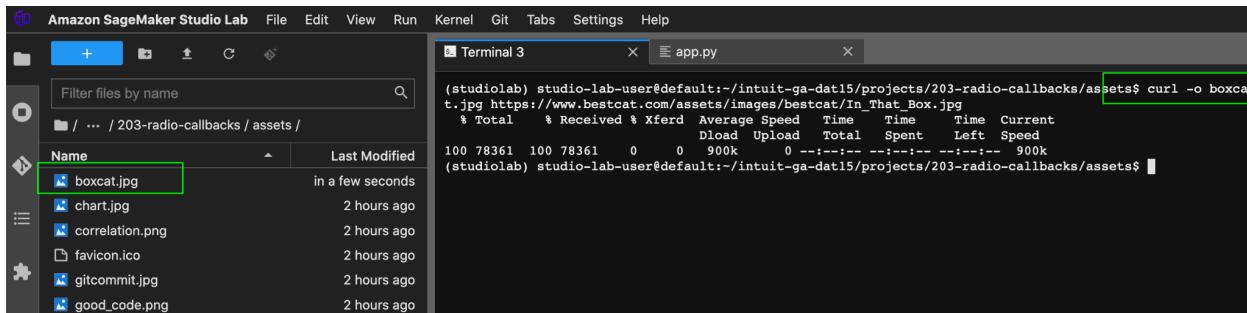
There's a file in here called "favicon" which is the tiny image that displays in your browser tab. You can find cool new favicons on this website: <https://www.favicon.cc/>

The other jpg and png files correspond to our "list of images" in line 13 of app.py - let's try adding a new file here. Sagemaker has an option to upload files from your laptop:



But the "pro" way to do this is in terminal using curl. Make sure you're inside the "assets" folder when you do this.

```
curl -o boxcat.jpg https://www.bestcat.com/assets/images/bestcat/In_That_Box.jpg
```



Now modify "app.py" to use the new file. Make sure that the name of the image in line 13 matches what's in your "assets" folder exactly. Notice that down on line 33, we are using indexing to call the values from our lists - no hard coded values here! - so you don't have to update this line (unless you end up changing the order or number of images in the list).

```

8 ##### Define your variables #####
9
10 myheading1='How to use callbacks'
11 tabtitle = 'xkcd'
12 list_of_options=['cat in a box!', 'correlation', 'git commit', 'scatterplot']
13 list_of_images=['boxcat.jpg', 'correlation.png', 'gitcommit.jpg', 'scatterplot.png', 'good_code.png']
14 sourceurl = 'https://xkcd.com/'
15 githublink = 'https://github.com/plotly-dash-apps/203-radio-callbacks'
16
17 ###### Set up the chart
18
19 ###### Initiate the app
20 external_stylesheets = ['https://codepen.io/chiddyp/pen/bWLwgP.css']
21 app = dash.Dash(__name__, external_stylesheets=external_stylesheets)
22 server = app.server
23 app.title=tabtitle
24
25 ###### Set up the layout
26
27 app.layout = html.Div(children=[
28     html.H1(myheading1),
29     dcc.RadioItems(
30         id='your_input_here',
31         options=[{
32             'label':list_of_options[0], 'value':list_of_images[0]},
33             {'label':list_of_options[1], 'value':list_of_images[1]},
34             {'label':list_of_options[2], 'value':list_of_images[2]},
35             {'label':list_of_options[3], 'value':list_of_images[3]},
36             ],
37             value=list_of_images[4],
38         ),
39         html.Div(id='your_output_here', children=''),
40     ]
41 )

```

Now we repeat all the same steps as before.

9	Test out the file “app.py”	python app.py
10	Check to see that the following message is displayed in terminal:	Dash is running on http://127.0.0.1:8050/
11	Stop running the file “app.py”	keyboard: ctrl+c
12	Add your changes to git	git add .
13	Write a commit message	git commit -m "I added a picture of a cat in a box!"
14	Push your changes to the remote origin	git push

At this point, github will ask you for your username and your “password” (which is actually your personal access token).

Check out the remote origin of your repo to make sure that the changes were successful. Over in heroku, go to the “Deploy” tab, scroll down to the bottom, and choose “Deploy Branch”. If everything went smoothly, your app will update with your changes! Note that sometimes the browser on your laptop will cache the previous results - so you may need to open the app in a different browser or in “incognito mode” in order to see the latest version.

When you’re done, post the URL of your app in the #peer-sharing channel so we all celebrate your success!

7. Final thoughts

A lot of you are probably wondering why we are spending so much time on deploying python apps with cat pictures - this is a class about data science and machine learning, right? We need to lay a foundation for deploying simple apps using Sagemaker, Plotly Dash and Heroku. Once you get down the pattern with simple apps, you will be able to build increasingly complex applications and incorporate real machine learning models - but the basic pattern (build, push, deploy) remains the same. Every week during the final hour of class we'll do this same pattern with a new example until it becomes second nature.

In fact, the pattern of deploying to Heroku via github will actually become pretty tedious very quickly - it's a nuisance to have to use github as a middle-man. In addition, it's not possible to view your changes (or your mistakes) until going through this same long process each time. Unfortunately, at this time Sagemaker does not support two key tools that developers rely on:

- Downloading and installing the Heroku CLI tool (so you can skip github completely)
- Viewing your app on localhost before you push and deploy

It's outside the scope of our class to do everything, and we really need to have Sagemaker because it provides so many powerful tools for training and testing high-power machine learning models - something that is difficult to do with your typical laptop. So for the purpose of the class, I will continue using only Sagemaker for demos.

However, if you're feeling confident in your skillset, you may choose to start developing your apps on your laptop, even while you continue to train your models in Sagemaker. This will allow you to install the Heroku CLI and also to view your app on localhost prior to deploying. It's up to you (this is not required, and it's also not recommended for those with limited experience).

You can download and install Python to your laptop using Anaconda like this:
<https://www.anaconda.com/products/individual>

You can download and install the Heroku CLI to your laptop like this:
<https://devcenter.heroku.com/articles/heroku-cli>

Here's a blog post I wrote, showing how to do all the steps from today's project using your laptop and the CLI.

<https://austinlasseter.medium.com/deploy-a-plotly-dash-app-on-heroku-4d2c3224230>

Good luck!