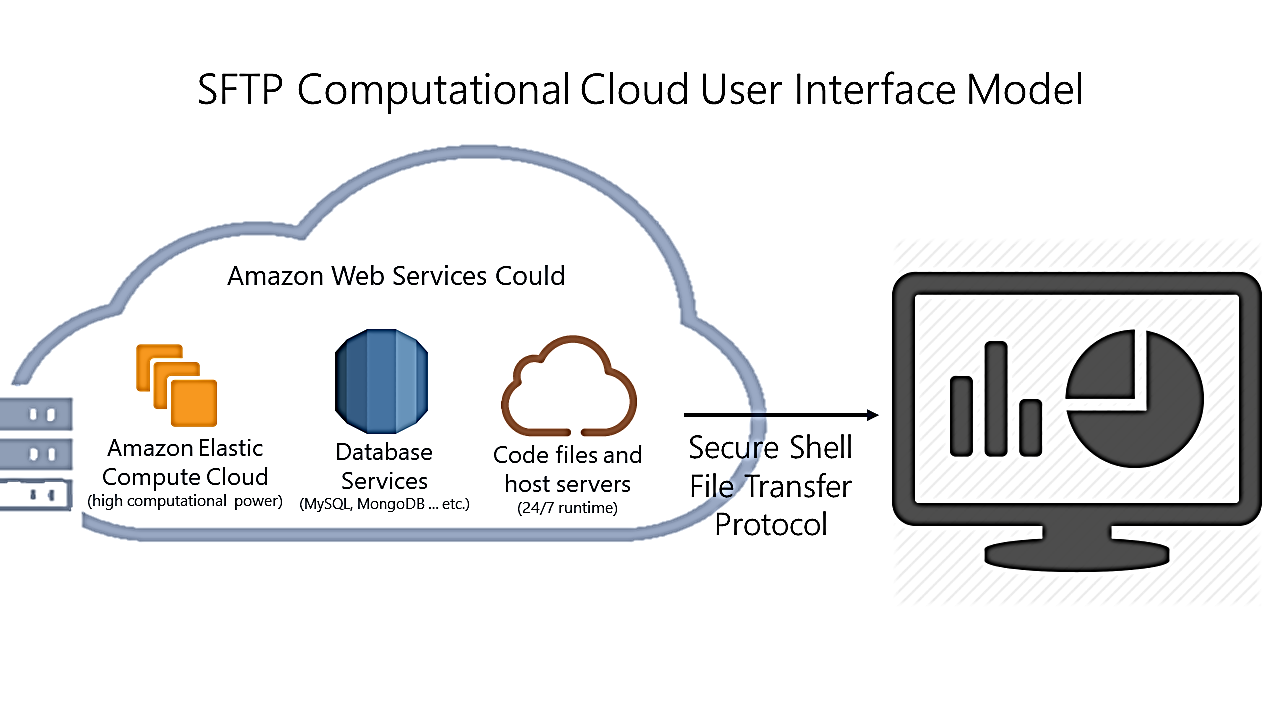
Development Environment Setup Guide

Computational and Storage Facilities

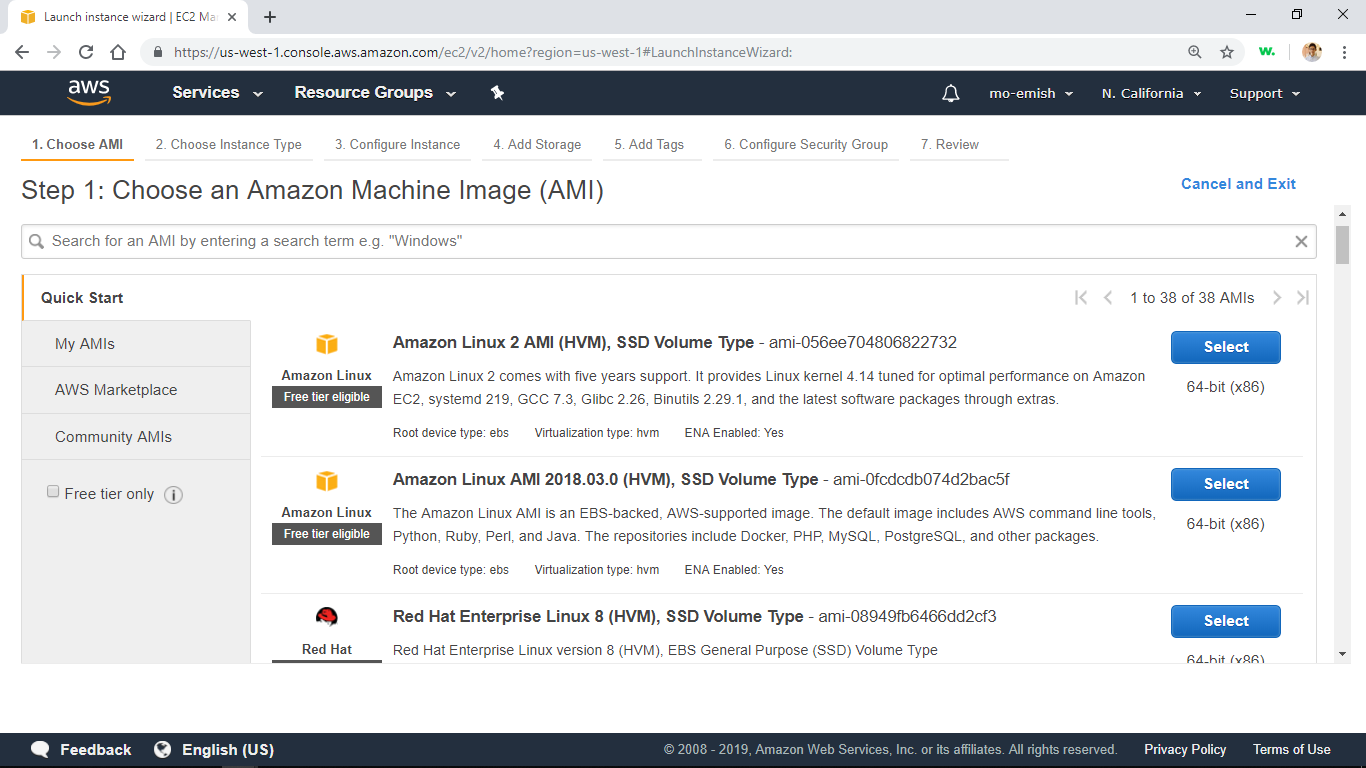
Amazon Web Services

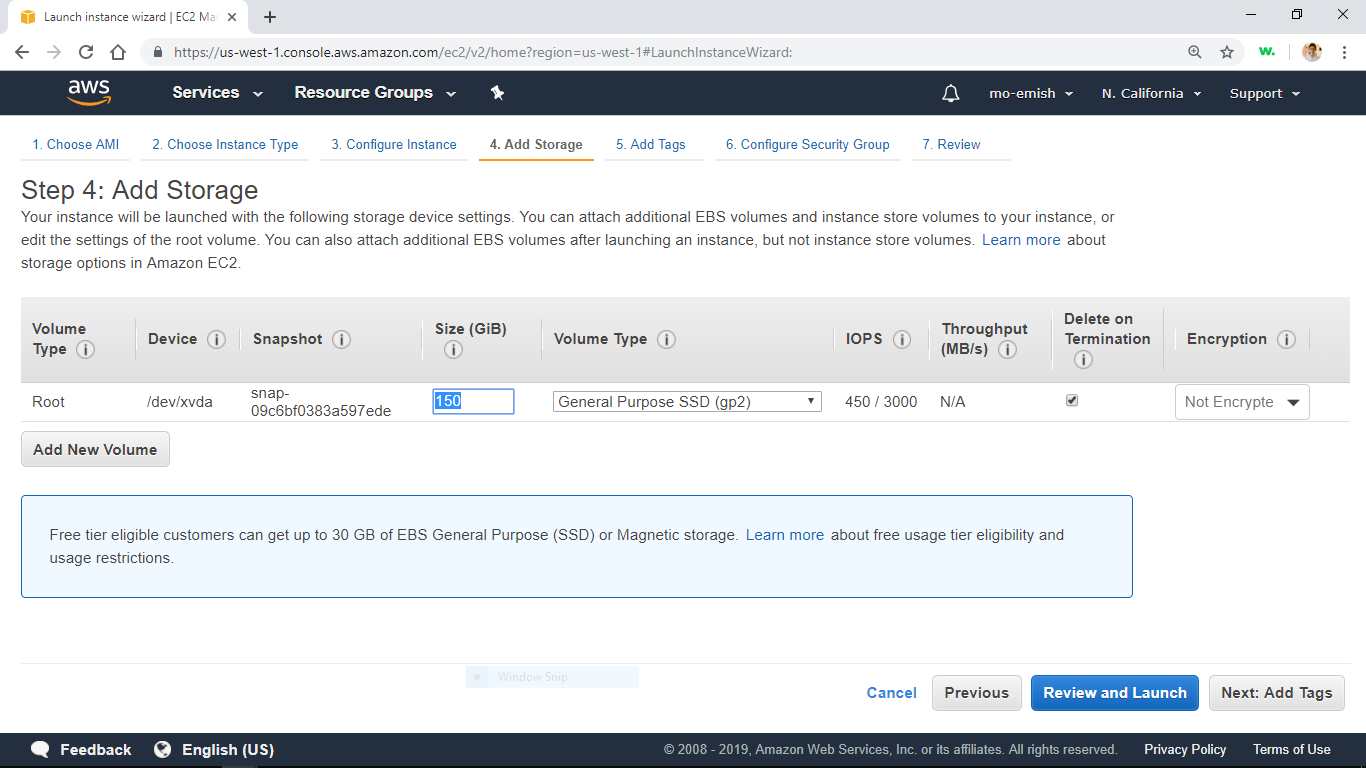
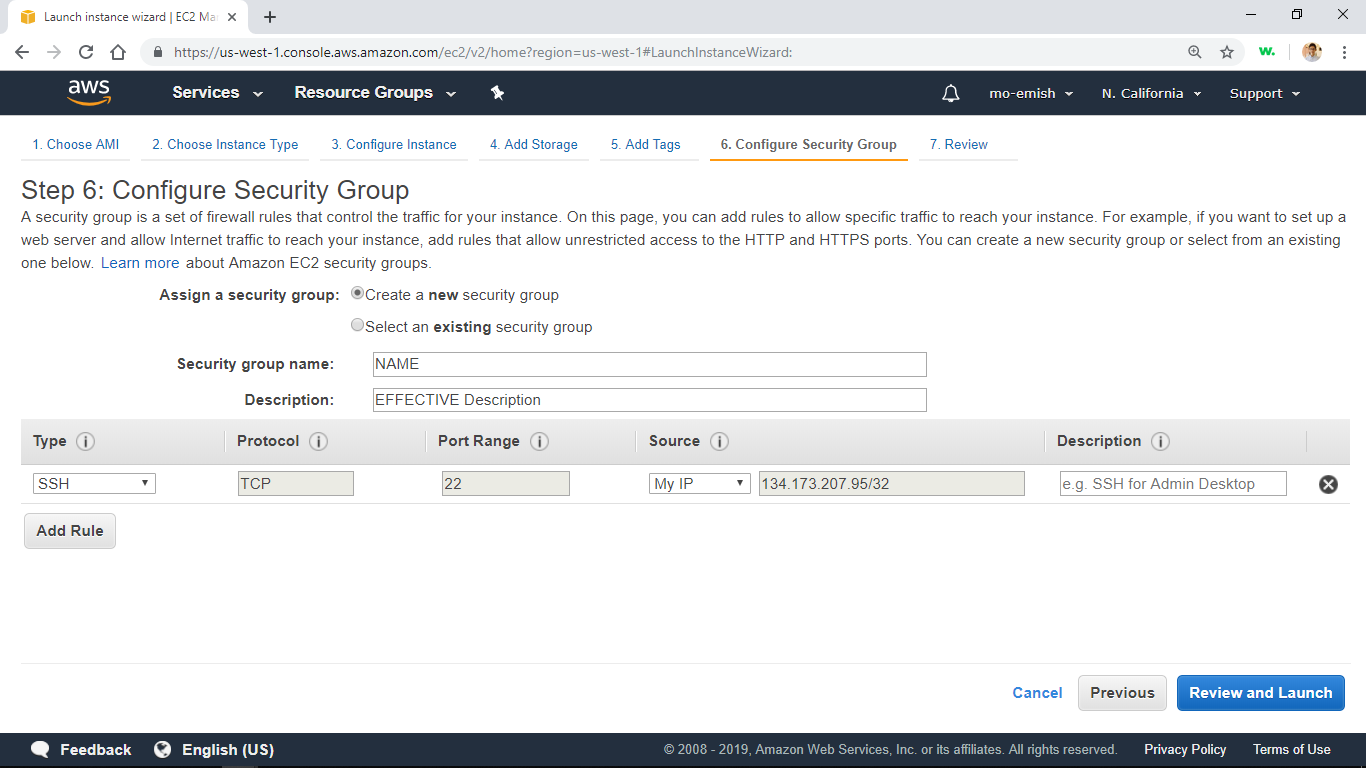
Amazon Web Services (AWS) is a cloud computing and storage services that was chosen to host the development environment about to be presented. AWS offers high-end computational power and compatibility with mass-scale used softwares and technology solutions. The following diagrams presents more information on the features and advantages of using AWS.

Amazon Elastic Compute Cloud

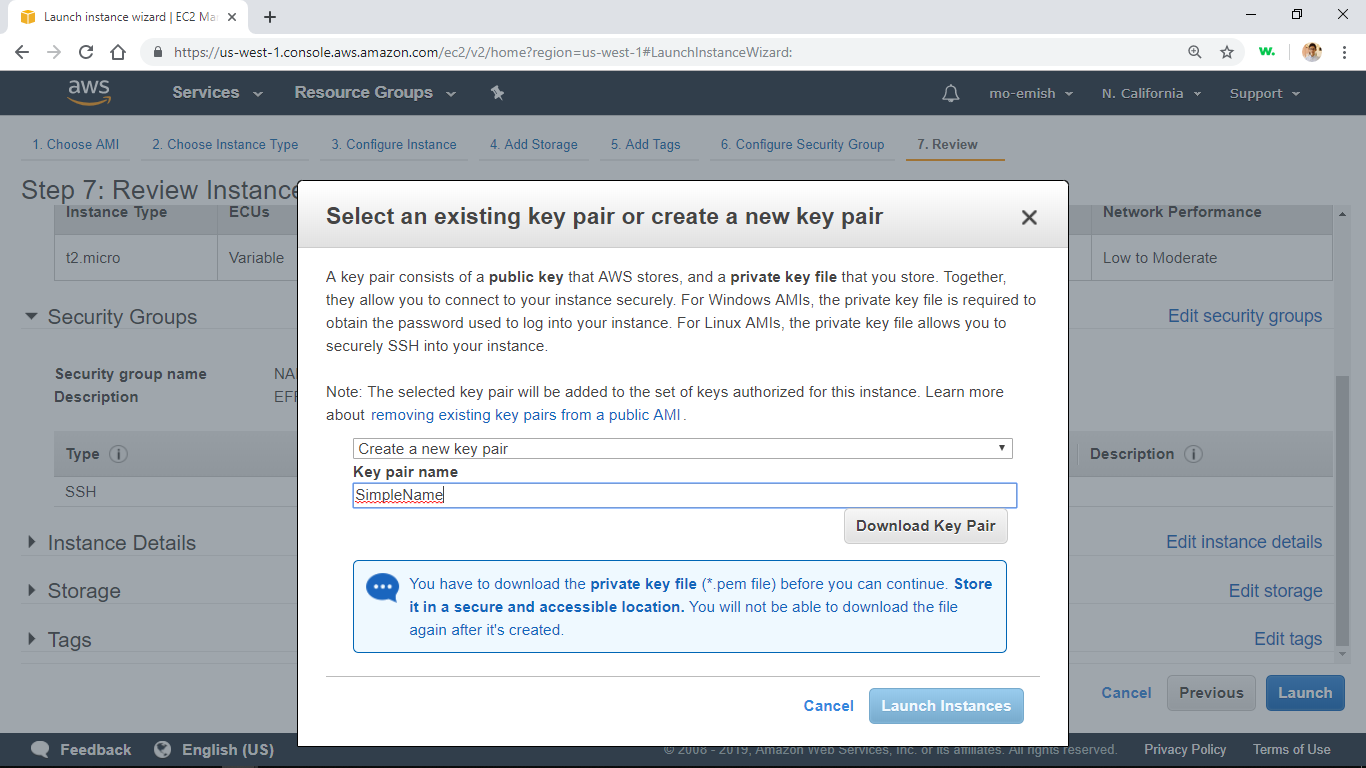
Amazon Elastic Compute Cloud, or EC2, is one of the services offered on AWS. EC2 is an on-demand computational service billed on hourly basis. The aim of using EC2 is creating a virtual machine that will run the codes and store code files as well as big data files. To create a virtual machine on EC2:

1. Go to [http://aws.amazon.com](http://aws.amazon.com/) and create an account
2. Under *services* tab at the top of the counsel homepage look for EC2 and press on it
3. Press the blue button that says *Launch Instance*
4. Next you will be asked to choose the operating system. For our purposes, please choose *Amazon Linux AMI 2018.03.0 (HVM)* as shown in the figure below



1. You will be asked to select the specifications of the machine. For our purposes, we chose t3.xlarge machine with the following specifications (4 processing cores, 16 GBs of ram) Choose a machine and press the white button that says *Next: Configure Instance Detials*
2. Press *Next: add storage*
3. Choose storage amount for the virtual machine. We used 150 GB as shown below. Choose storage and press *Next: add tags*
4. Press *Next: Configure Security Groups*
5. Under *Source* column in the table choose *My IP* to make sure the virtual machine can only be accessed from your computer. You can add more machines later if you choose to. Name the security group and give it a description as show bellow. After you are done press the blue button *Review and Launch* 

Note: in case you are intending to use RStudio, press add rule, enter 8787 in port, and set source to My IP.

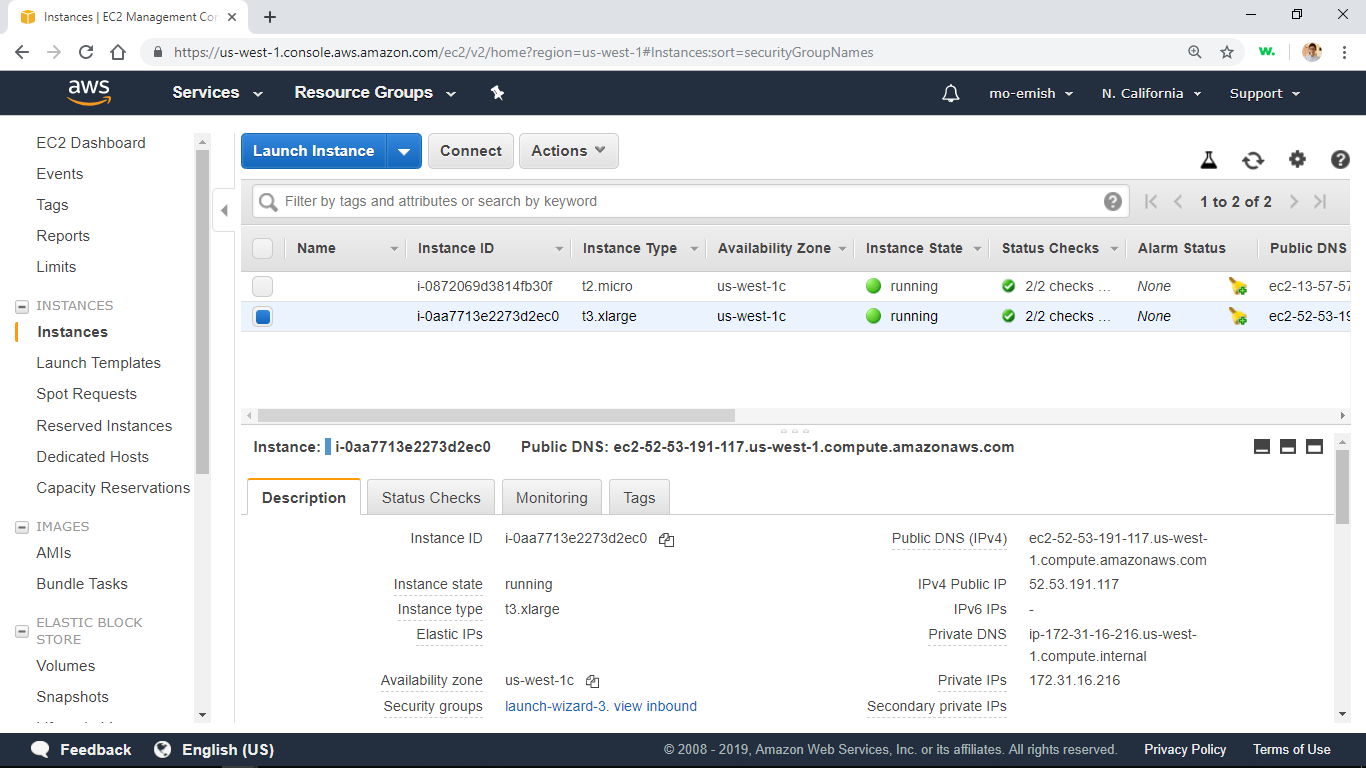
1. Press *Launch.* You will then be asked to choose key pair. Choose create new key pair, give your key pair a simple name, download it, and then press *Launch Instance.* Figure bellow.Note that key pair is an important file that allows you to access the virtual machine you just created. Make sure its name is simple, it is stored in a place that is easy to access, and always keep a copy of it. 

Now in order to access the virtual machine on your computer:

1. Change the key pair permissions so that only the current used has full control over it. This step will vary based on which operating system you are using.
2. Open Command Prompt on windows, or Terminal on Mac
3. Type in the following command:

ssh –i path\to\your\key\pair.pem ec2-user@machine’s Public DNS

make sure to replace the red parts with the parts they correspond to

1. To get your machines public DNS go to AWS counsel, EC2, and press view instances. Press the box next to the machine you are trying to connect to and then you will find public DNS under the description tab as shown. 
2. Upon successful connection you will see the following message

Last login:DATE from IP

\_\_| \_\_|\_ )

\_| ( / Amazon Linux AMI

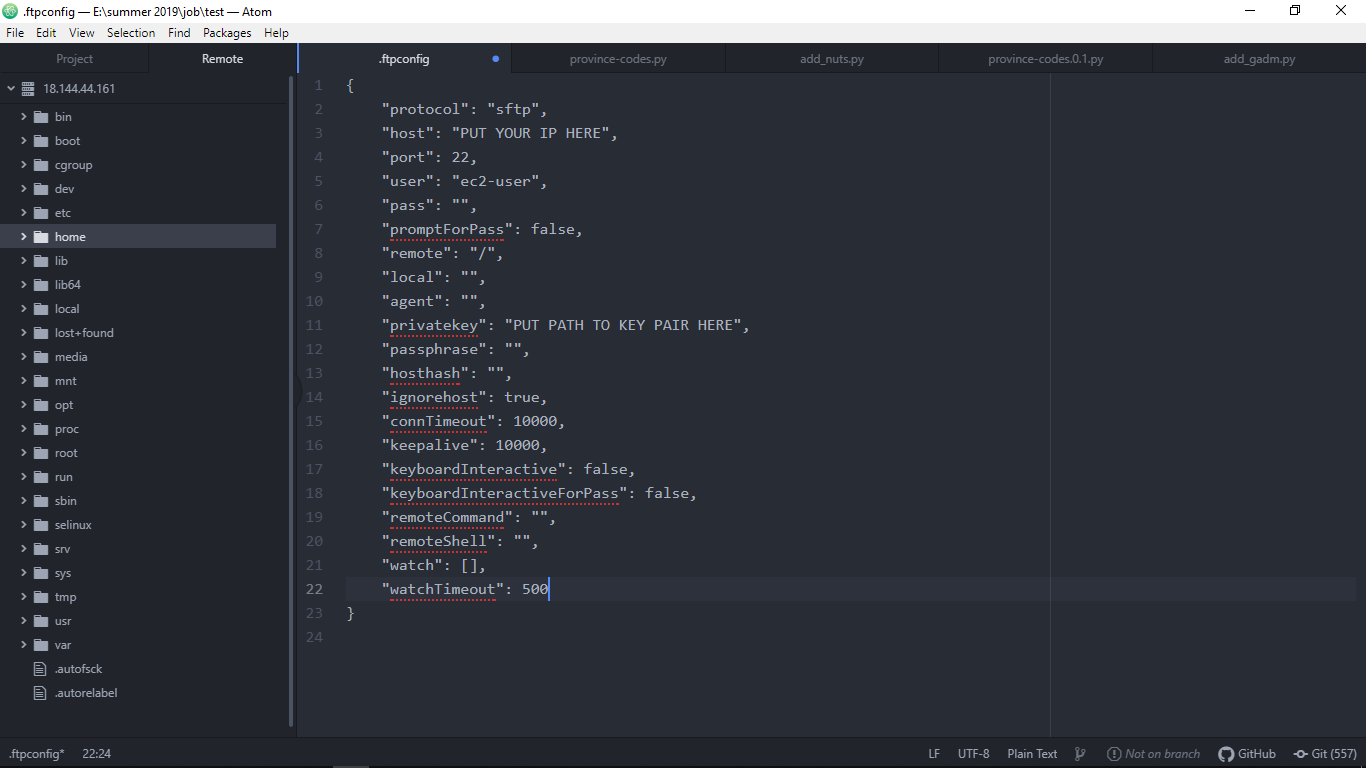
\_\_\_|\\_\_\_|\_\_\_|

Logistic solutions

In order to facilitate running commands and accessing files on the EC2 instance, we recommend using the following softwares:

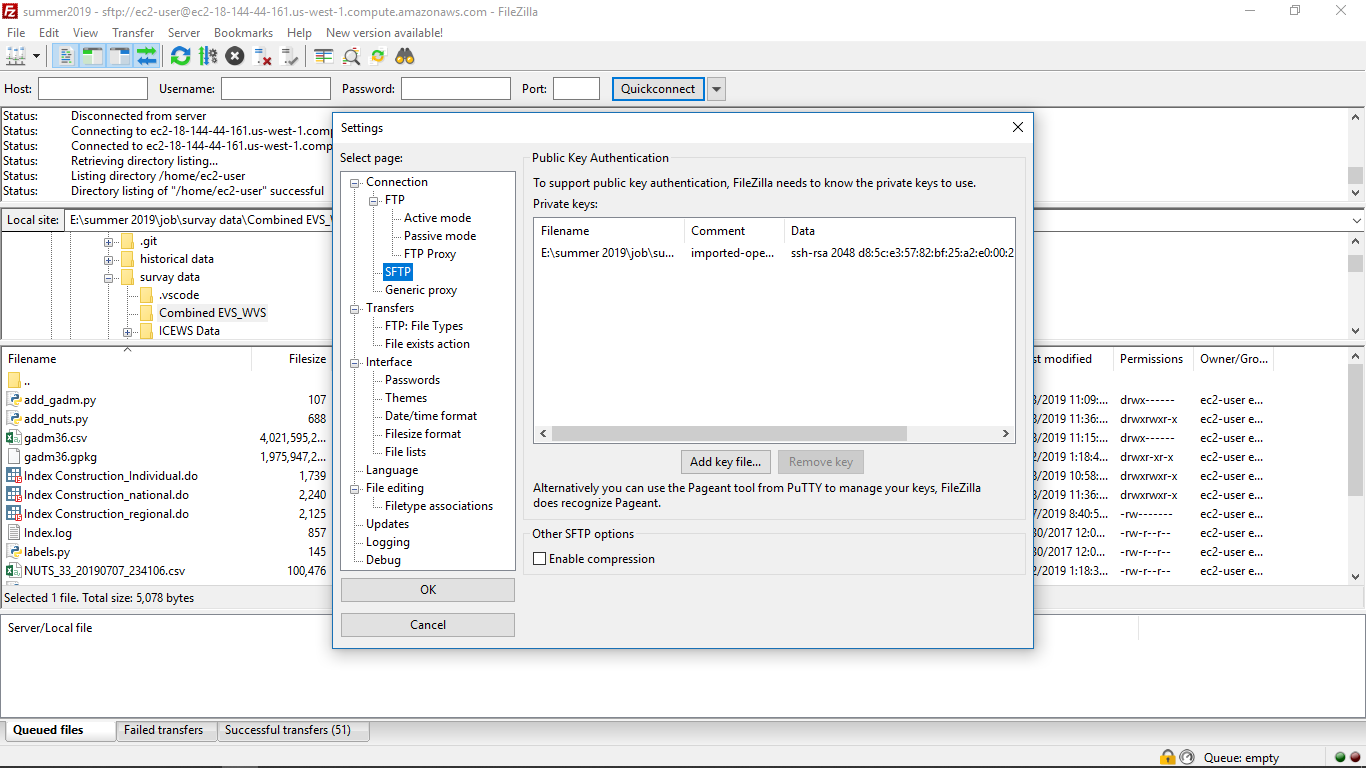
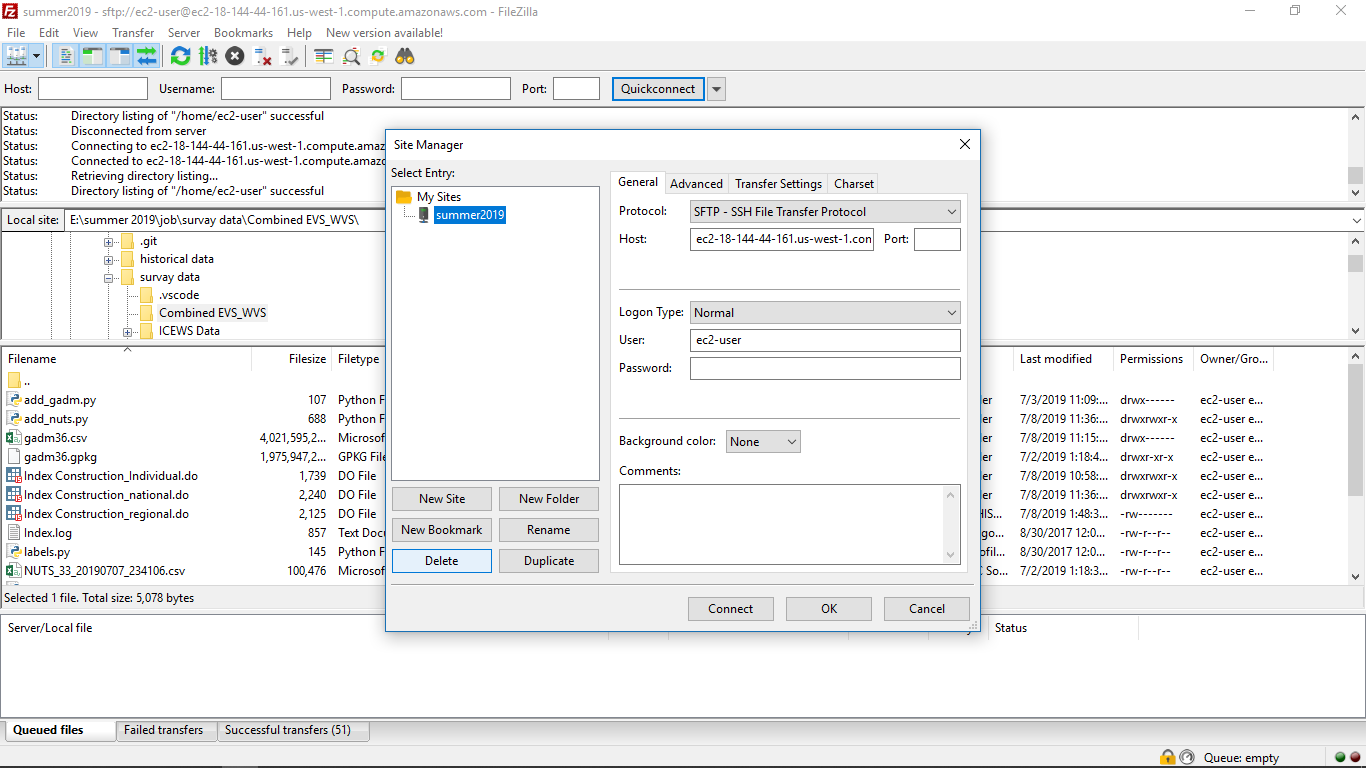
Atom

Text editor that synchronizes and facilitates accessing code files. To connect Atom to EC2 instance:

1. Install Atom on your computer: <https://atom.io/>
2. Install Remote FTP package on Atom: <https://atom.io/packages/remote-ftp>
3. Under the packages menu in atom choose Create SFTP configuration file and fill in the file as follows 
4. Make sure to save configuration file
5. From packages menu, hover over Remote FTP again and click toggle
6. Click on remote tab in the sidebar and then click connect. Now you can access the files in the sidebar

FileZilla

FileZilla is a file browser with a friendly graphical user interface that allow easy file access and transfer. To setup FileZilla:

1. Install FileZilla Free client version   
   <https://filezilla-project.org/download.php?type=client>
2. Open the edit menu and click settings
3. Choose STFP on the left side and then import your key pair 
4. Open the sites manager (the button under file menu) and press create new site
5. Rename the site in the left sidebar, choose *SFTP* from protocol dropdown menu, add your machine’s public DNS as host, and set user name to *ec2-user* 
6. Press connect and now you can see the contents of your EC2 machine on the right side of the page

Required Software

STATA

To install STATA 15 on EC2 instance:

1. Download STATA for Linux in .tar.gz format
2. In the same file where the .tar.gz file exists type the command: *tar xvzf file.tar.gz*
3. Make the installation directory */usr/local/stata15* and cd into it on the command line.
4. Type */path/to/extracted/files/install*
5. Follow instructions and make sure to activate the license after installing (type ./statainit and follow instructions)
6. In order to add the installation directory to the system path, type the following commands:  
   echo export PATH="/usr/local/stata14:$PATH" >> ~/.bashrc  
   source ~/.bashrc
7. To run STATA do files type: *stata do* *filename.do*

Python 3.7, pip, and other packages

To install python 3.7 on EC2 instance:

1. Type *sudo yum install python3*

After Python 3.7 is successfully installed, proceed to install pip using the following steps:

1. Type *curl -O https://bootstrap.pypa.io/get-pip.py*
2. Type *python3 get-pip.py –user*
3. Verify pip is installed properly by typing *pip –version*

Now, use pip to install the python packages used in code file. The following table shows the commands used to install the required libraries.

|  |  |
| --- | --- |
| Pandas | pip install pandas |
| Numpy | pip install numpy |
| Country\_converter | pip install country\_converter |
| pyreadstat | pip install pyreadstat |
| iso3166 | pip install iso3166 |
| pycountry | pip install pycountry |

In order to run the python files type: *python3 filename.py*

R & RStudio

In order to install R type the following command into the EC2 machine: *sudo yum install -y R*

Next install RStudio using the following steps:

1. Type *wget https://download2.rstudio.org/rstudio-server-rhel-1.0.153-x86\_64.rpm*
2. After the file is downloaded type   
   *sudo yum install -y --nogpgcheck rstudio-server-rhel-1.0.153-x86\_64.rpm*
3. Make sure you have added a rule for port 8787 in step 9 of the instance setup in order to have access over RStudio from your browser
4. Go to the browser (e.g. Google Chrome) and type the machine’s Public DNS followed by :8787

MongoDB

To install MongoDB:

1. Download all .rpm files in <https://repo.mongodb.org/yum/amazon/2013.03/mongodb-org/4.2/x86_64/RPMS/>
2. Transfer files using fileZila
3. Type *sudo yum install filename.rpm* for each of these files