

---

# PREDICT 420

Atef Bader, PhD

# Agenda

---

- SSCC setup for dornick & postgres
- Assignment #1 Walkthrough & Deliverable
- Exercise #3 Walkthrough & Deliverable

---

# VPN & SSCC Setup

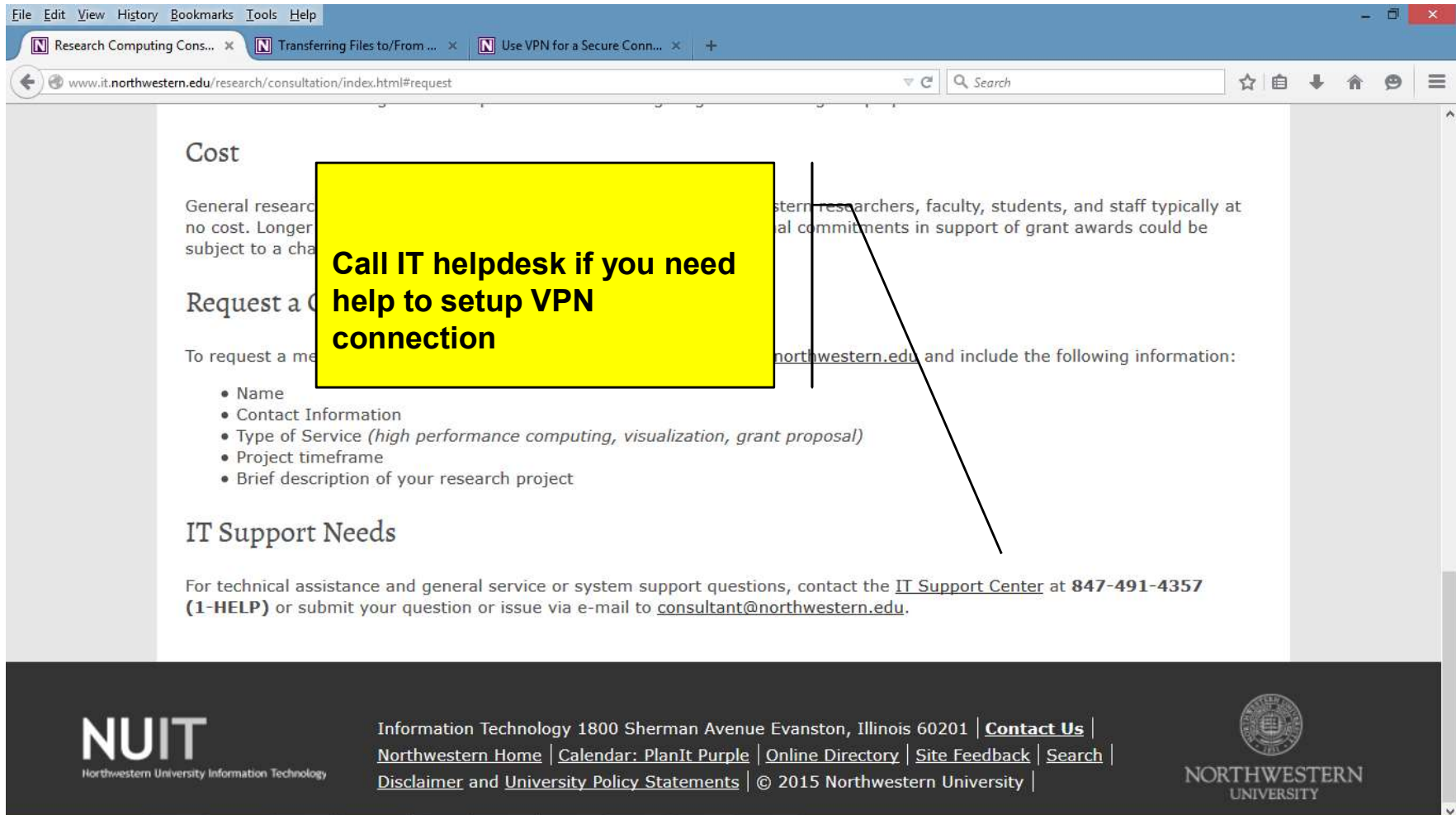
# Be Patient ...

---



- It will take you sometime (more or less 2 hours) to complete the setup for VPN to dornick and postgres

# CALL HELPDESK if you need help with VPN



The image is a screenshot of a web browser displaying the Northwestern University IT Support Center website. The browser's address bar shows the URL [www.it.northwestern.edu/research/consultation/index.html#request](http://www.it.northwestern.edu/research/consultation/index.html#request). The page content includes sections for 'Cost', 'Request a Consultation', and 'IT Support Needs'. A prominent yellow rectangular callout box with a black border is overlaid on the page, containing the text 'Call IT helpdesk if you need help to setup VPN connection'. A black line points from this callout box to the 'IT Support Needs' section. The footer of the page contains the NUIT logo, contact information, and various links.

**Cost**

General research... no cost. Longer... subject to a cha...

**Request a Consultation**

To request a me... northwestern.edu and include the following information:


- Name
- Contact Information
- Type of Service (*high performance computing, visualization, grant proposal*)
- Project timeframe
- Brief description of your research project

**IT Support Needs**

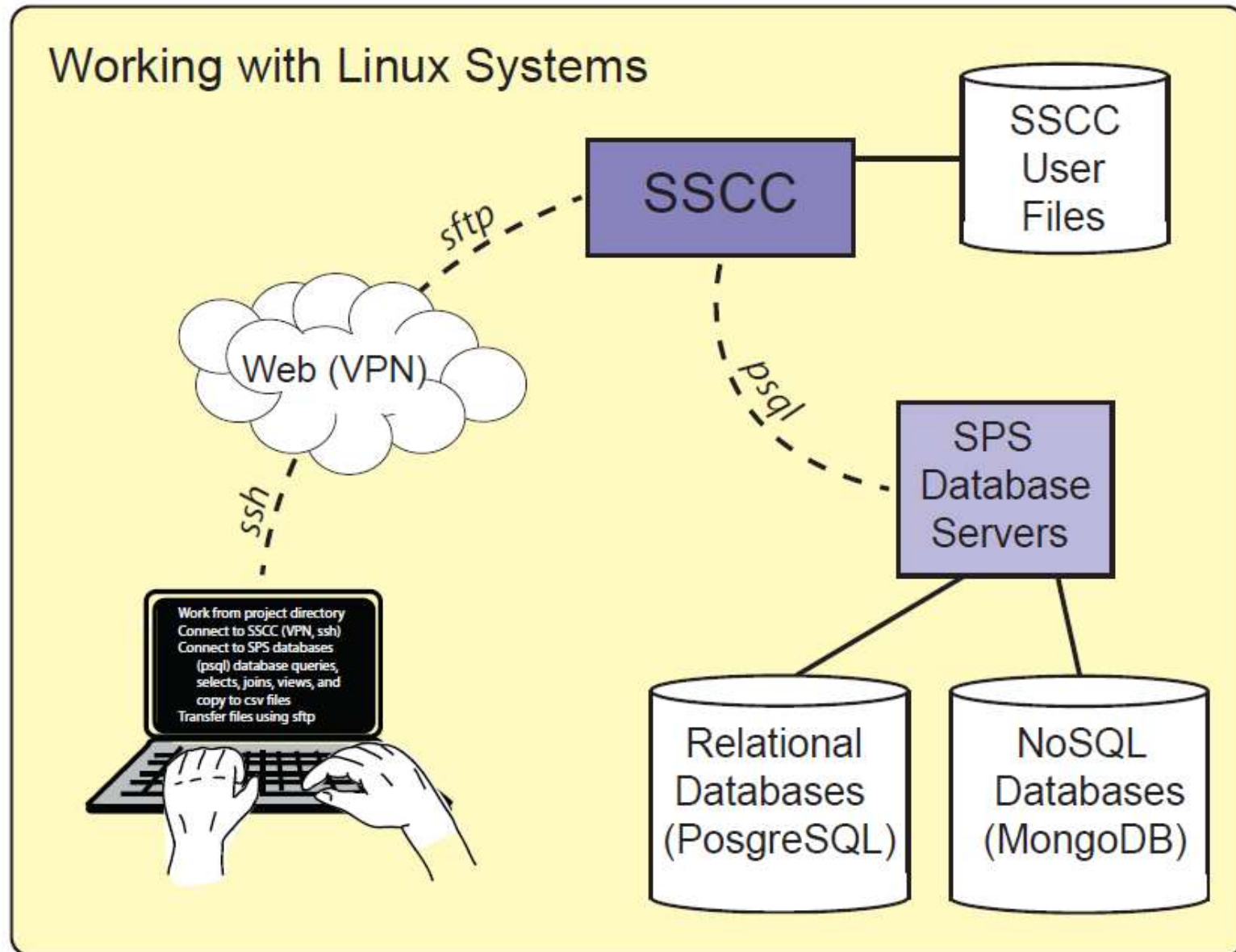
For technical assistance and general service or system support questions, contact the [IT Support Center](#) at **847-491-4357 (1-HELP)** or submit your question or issue via e-mail to [consultant@northwestern.edu](mailto:consultant@northwestern.edu).

**NUIT**  
Northwestern University Information Technology

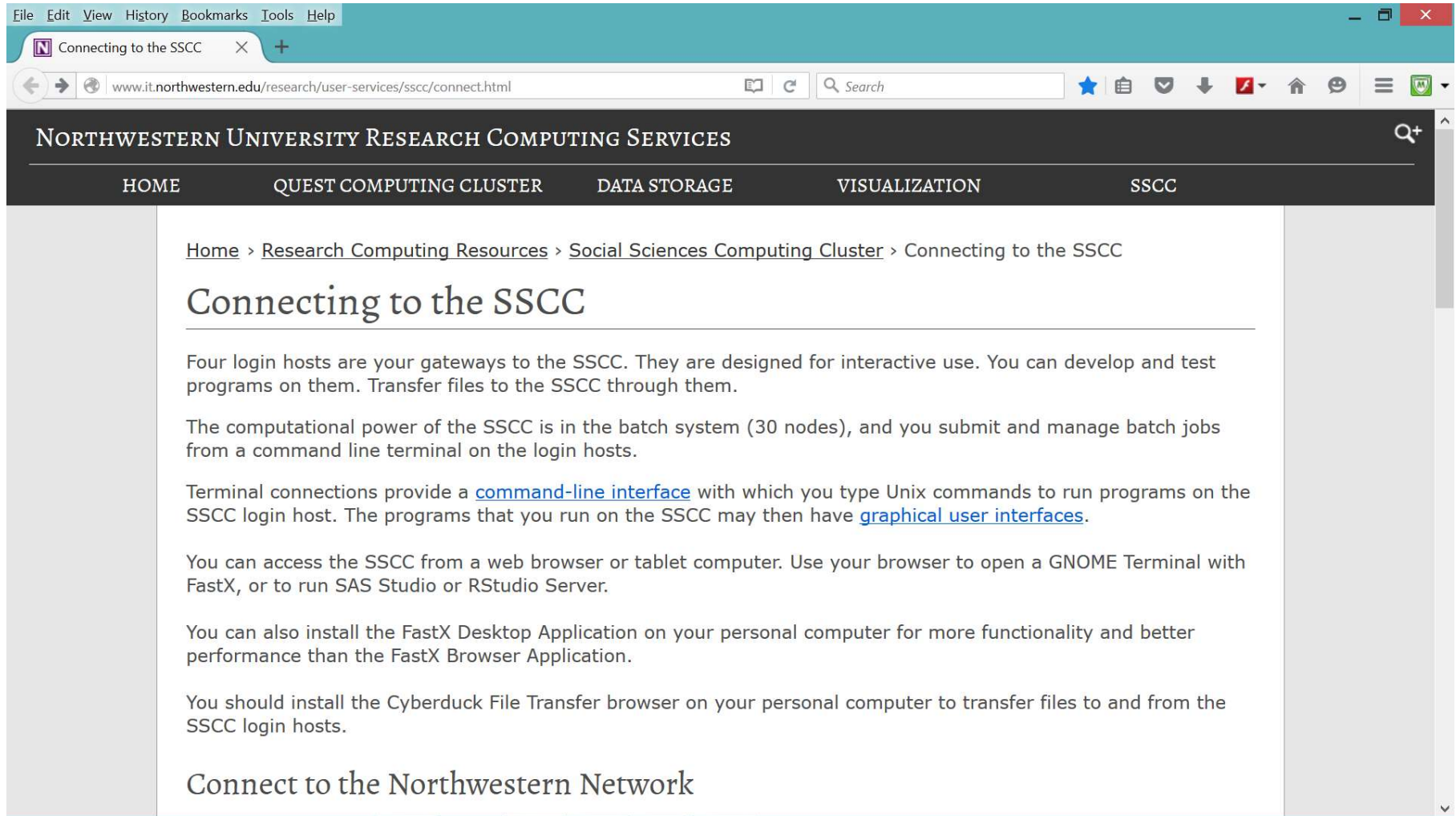
Information Technology 1800 Sherman Avenue Evanston, Illinois 60201 | [Contact Us](#) | [Northwestern Home](#) | [Calendar: PlanIt Purple](#) | [Online Directory](#) | [Site Feedback](#) | [Search](#) | [Disclaimer](#) and [University Policy Statements](#) | © 2015 Northwestern University

  
**NORTHWESTERN**  
UNIVERSITY

# Your Connection Path



# Bookmark this URL – What is SSCC



The screenshot shows a web browser window with the address bar displaying `www.it.northwestern.edu/research/user-services/sscc/connect.html`. The page title is "Connecting to the SSCC". The header of the website is "NORTHWESTERN UNIVERSITY RESEARCH COMPUTING SERVICES" with a navigation menu containing "HOME", "QUEST COMPUTING CLUSTER", "DATA STORAGE", "VISUALIZATION", and "SSCC". The main content area has a breadcrumb trail: [Home](#) > [Research Computing Resources](#) > [Social Sciences Computing Cluster](#) > Connecting to the SSCC. The main heading is "Connecting to the SSCC". The text explains that four login hosts are gateways to the SSCC for interactive use, and that the computational power is in the batch system (30 nodes) accessible via command-line terminal. It also mentions terminal connections providing a [command-line interface](#) and programs having [graphical user interfaces](#). Further, it states that the SSCC can be accessed from a web browser or tablet computer using a GNOME Terminal with FastX, or SAS Studio or RStudio Server. It also notes that the FastX Desktop Application can be installed on a personal computer for better performance, and that the Cyberduck File Transfer browser should be installed for file transfer.

Home > Research Computing Resources > Social Sciences Computing Cluster > Connecting to the SSCC

## Connecting to the SSCC

Four login hosts are your gateways to the SSCC. They are designed for interactive use. You can develop and test programs on them. Transfer files to the SSCC through them.

The computational power of the SSCC is in the batch system (30 nodes), and you submit and manage batch jobs from a command line terminal on the login hosts.

Terminal connections provide a [command-line interface](#) with which you type Unix commands to run programs on the SSCC login host. The programs that you run on the SSCC may then have [graphical user interfaces](#).

You can access the SSCC from a web browser or tablet computer. Use your browser to open a GNOME Terminal with FastX, or to run SAS Studio or RStudio Server.

You can also install the FastX Desktop Application on your personal computer for more functionality and better performance than the FastX Browser Application.

You should install the Cyberduck File Transfer browser on your personal computer to transfer files to and from the SSCC login hosts.

### Connect to the Northwestern Network

# Bookmark this URL – Creating VPN

File Edit View History Bookmarks Tools Help

Connecting to the SSCC Use VPN for a Secure Con...

www.it.northwestern.edu/oncampus/vpn/index.html

**NUIT**  
Northwestern University Information Technology

Information Technology Northwestern University

IT KNOWLEDGE BASE SUPPORT SERVICES LEARN/TEACH RESEARCH COMPUTING ABOUT NUIT  
SYSTEM LOGINS

**Additional Information**

**Use VPN for a Secure Connection**

- Knowledge Base VPN Technical Support
- SSL VPN Overview
- Staying Secure During Travel

**Get Connected**

**NUIT Home > Get Connected > VPN**

## Use VPN for a Secure Connection

**Chrome Users:** Chrome browser version 42 is not compatible with VPN. If you experience issues accessing this service, switch to another browser - Internet Explorer, Firefox, or Safari. [Learn more.](#)

Virtual Private Network (VPN) establishes a "secure tunnel" for your computer on the Northwestern network. It provides strong encryption, enables authenticated access to the Northwestern network from external and untrusted environments, and proxies your network traffic so it appears to originate from within the Northwestern network. A VPN connection is an extension of the Northwestern network and subject to the University's policies. VPN connections are automatically ended twelve hours after initiation for service and security reasons.

Some Northwestern departments have determined that it is in the University's best interest to require the security of VPN before accessing their resources from off-campus locations. Services that require connecting to Northwestern VPN before accessing them include:

- Kronos
- Remote desktop access
- NUFinancials

In addition, many servers and other resources (e.g., printers) managed by schools and departments



# Bookmark this URL – Connecting to SSCC

What laptop you have?





connecting to Northwestern VPN before accessing them include:

- Kronos
- Remote desktop access
- NUFInancials

In addition, many servers and other resources (e.g., printers) managed by schools and departments require VPN before accessing them. You may want to talk to your local technical support person before trying to access resources from off campus.

NOTE: In the event of a University-wide campus closure, instructions for accessing emergency VPN will be available for you.

### Set Up VPN

	Step-by-Step Walkthroughs	How-To Videos
<b>Mac OS X</b>	<a href="#">Yosemite (10.10), Mavericks (10.9) or Mountain Lion (10.8)</a>	<a href="#">Connect to Northwestern VPN Using Mac OS X</a> 
<b>Windows</b>	<a href="#">Windows 8.1</a>	<a href="#">Connect to Northwestern VPN Using Windows 8.1</a> 
	<a href="#">Windows 8</a>	<a href="#">Connect to Northwestern VPN Using Windows 8</a> 
	<a href="#">Windows 7</a>	<a href="#">Connect to Northwestern VPN Using Windows 7</a> 
<b>Mobile</b>	<a href="#">iPhone or iPad running iOS</a>	
	<a href="#">Android</a>	
	<a href="#">Windows Phone 8</a>	
<b>Linux</b>	<a href="#">Chrome OS</a>	
	<a href="#">Linux</a>	

Use Cisco AnyConnect VPN if you experience difficulty connecting with native VPN due to a firewall, NAT (Network Address Translation) environment, or a blocked VPN protocol due to service from an Internet service provider.

# Bookmark this URL – Connecting to SSCC

File Edit View History Bookmarks Tools Help

Connecting to the SSCC Use VPN for a Secure Con...

www.it.northwestern.edu/research/user-services/sscc/connect.html

nodes in the SSCC. No matter which SSCC computer you're using, you are

## Login Instructions by Intended Use

**Simple Command-Line:** Use the [FastX Browser Client](#) from your web browser, or the [FastX Desktop Application](#) installed on your personal computer. Login with FastX, and then *Choose* the **GNOME Terminal** bookmark. Suitable for editing files (use the command **evim**), creating and submitting batch jobs, and checking results.

**Command-Line with Full Graphics:** Use the [FastX Desktop Application](#). Choose the **GNOME Terminal** bookmark. Use this application for the Stata GUI by typing the command **xstata**, the MATLAB desktop using the command **matlab**, the RStudio Integrated Development Environment using the command **rstudio**, or the Python IDE using the commands **module load python; spyder**. See the SSCC [Analytical Software](#) page for a list of other IDEs.

**SAS Programming:** Use [SAS Studio](#) from your web browser.

**R Programming:** Use [RStudio Server](#) from your web browser, or login with the [FastX Desktop Application](#) and type the command **rstudio**.

## Transferring Files with Cyberduck

**Cyberduck** is a graphical program that runs on both Mac OS X and Microsoft Windows that functions as an SSH Secure Shell File Transfer application. It offers a conventional drag and drop interface. See [How to Transfer Files To/From the SSCC](#).

Last Updated: 14 January 2016

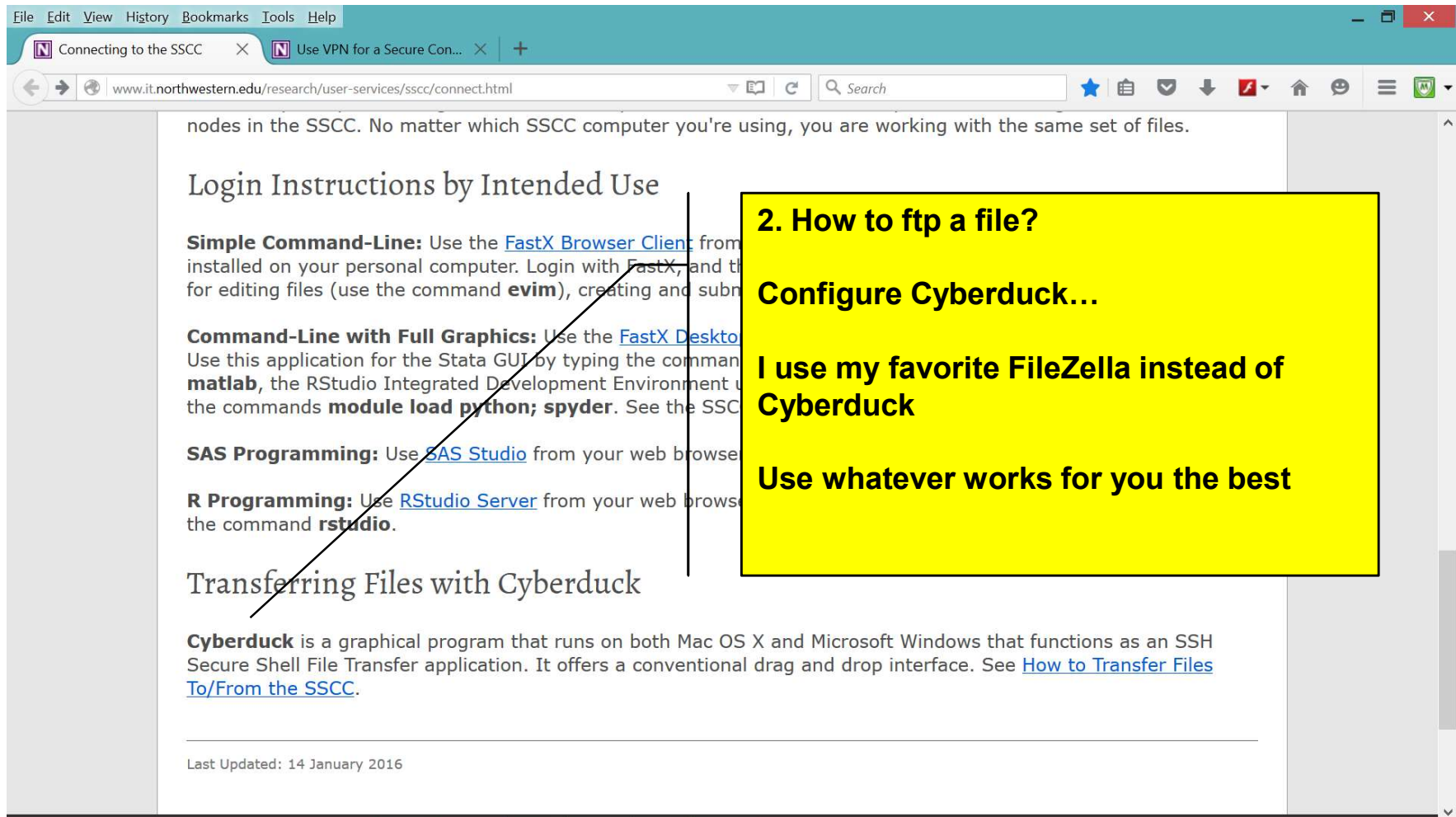
1. How to login?

Configure FastX...

I use my favorite Putty instead of FastX

Use whatever works for you the best

# Bookmark this URL – Connecting to SSCC



The screenshot shows a web browser window with the address bar displaying [www.it.northwestern.edu/research/user-services/sscc/connect.html](http://www.it.northwestern.edu/research/user-services/sscc/connect.html). The page content includes the heading "Login Instructions by Intended Use" and several sections: "Simple Command-Line", "Command-Line with Full Graphics", "SAS Programming", and "R Programming". A yellow callout box is overlaid on the right side of the page, containing the following text:

2. How to ftp a file?  
Configure Cyberduck...  
I use my favorite FileZella instead of Cyberduck  
Use whatever works for you the best

The page also includes a section titled "Transferring Files with Cyberduck" and a footer that reads "Last Updated: 14 January 2016".

# PuTTY– Connecting to SSCC

I use PuTTY to connect to SSCC – dornick server

I prefer it over X-Win-32

You need to download it in case you prefer to use it



## Download PuTTY

PuTTY is an SSH and telnet client, developed originally by Simon Tatham for the Windows platform. PuTTY is open source software that is available with source code and is developed and supported by a group of volunteers.

You can download PuTTY [here](http://www.putty.org).

## Bitvise SSH Client

Bitvise SSH Client is an SSH and SFTP client for Windows. It is developed and supported professionally by Bitvise. The SSH Client is robust, easy to install, easy to use, and supports all features supported by PuTTY, as well as the following:

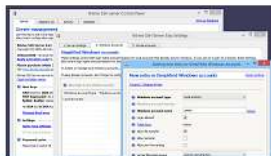
- graphical SFTP file transfer;
- single-click Remote Desktop tunneling;
- auto-reconnecting capability;
- dynamic port forwarding through an integrated proxy;
- an FTP-to-SFTP protocol bridge.

Bitvise SSH Client is **free for personal use**, as well as for individual commercial use inside organizations. You can [download Bitvise SSH Client here](http://www.bitvise.com/ssh-client.html).



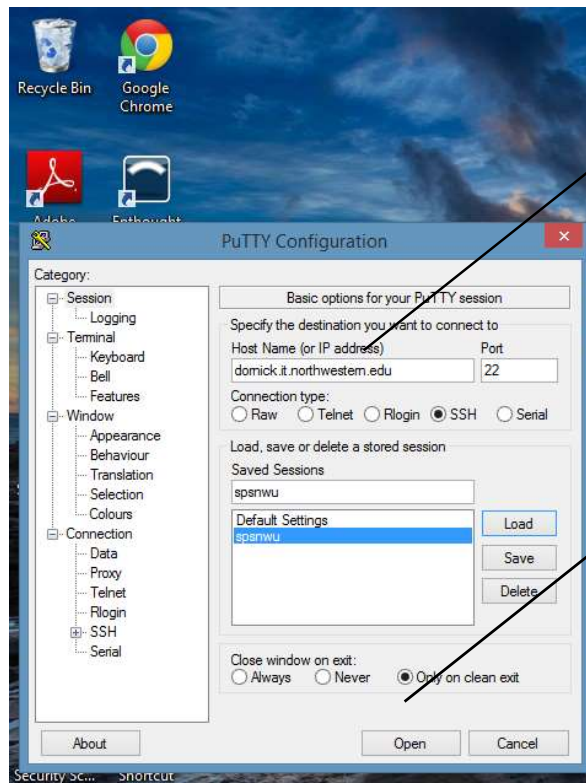
## Bitvise SSH Server

Bitvise SSH Server is an SSH, SFTP and SCP server for Windows. It is robust, easy to install, easy to use, and works well with a variety of SSH clients, including Bitvise SSH Client, OpenSSH, and PuTTY. The SSH Server is developed and supported professionally by Bitvise.





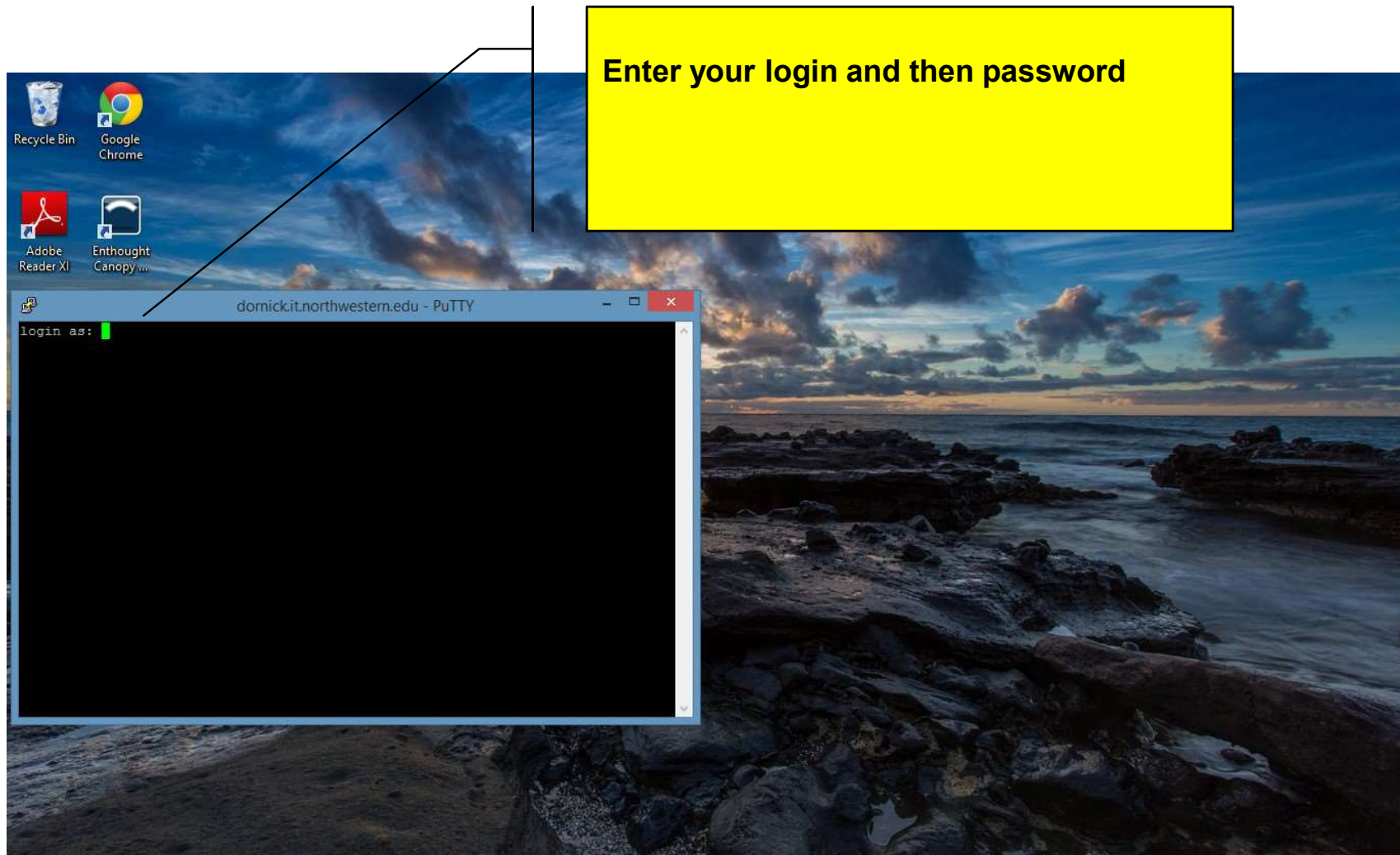
# PuTTY– Connecting to SSCC



Type dornick host

And then click Open button

# PuTTY– Connecting to SSCC



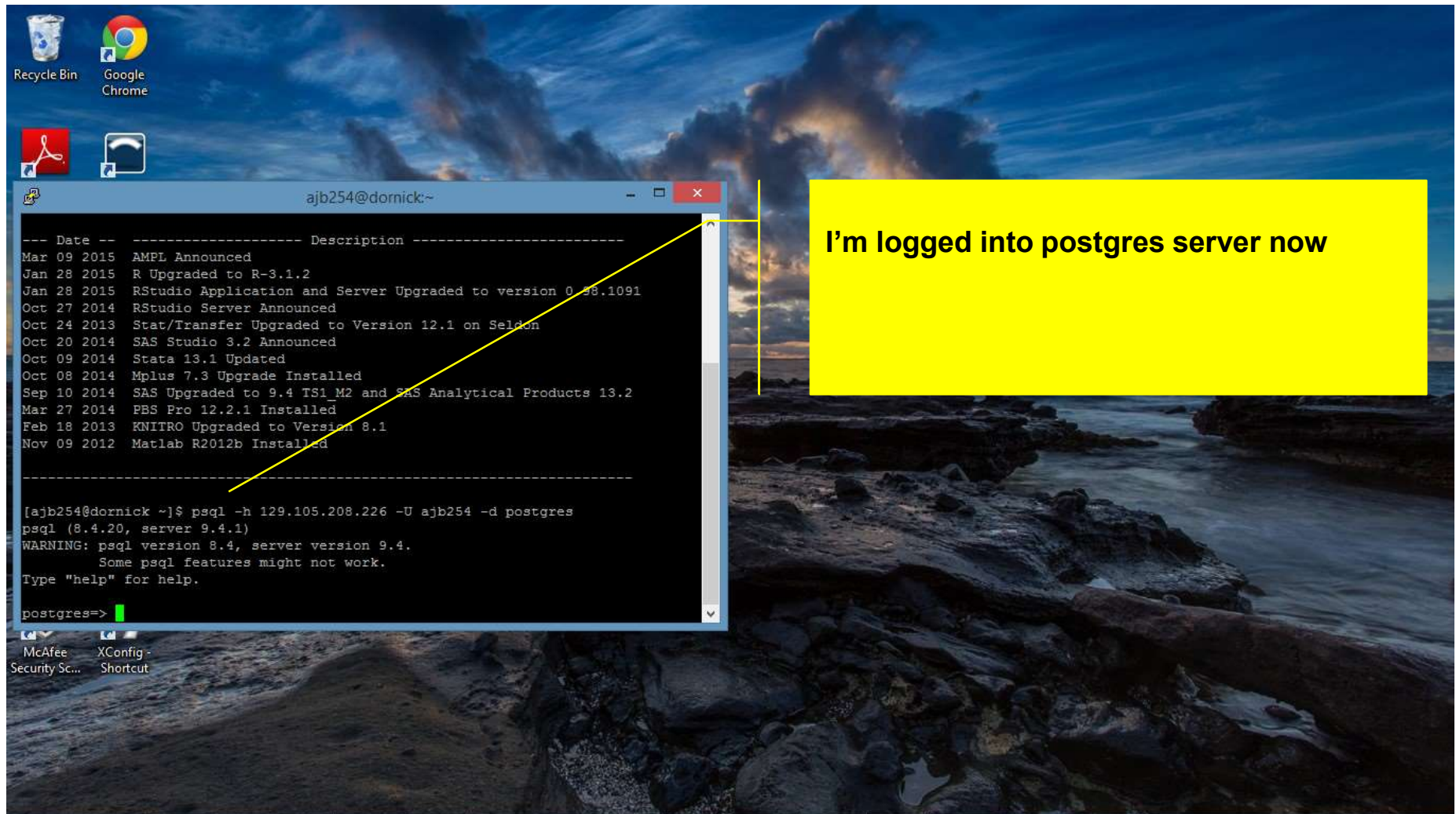
# PuTTY– Connecting to SSCC



I'm logged in now to dornick host

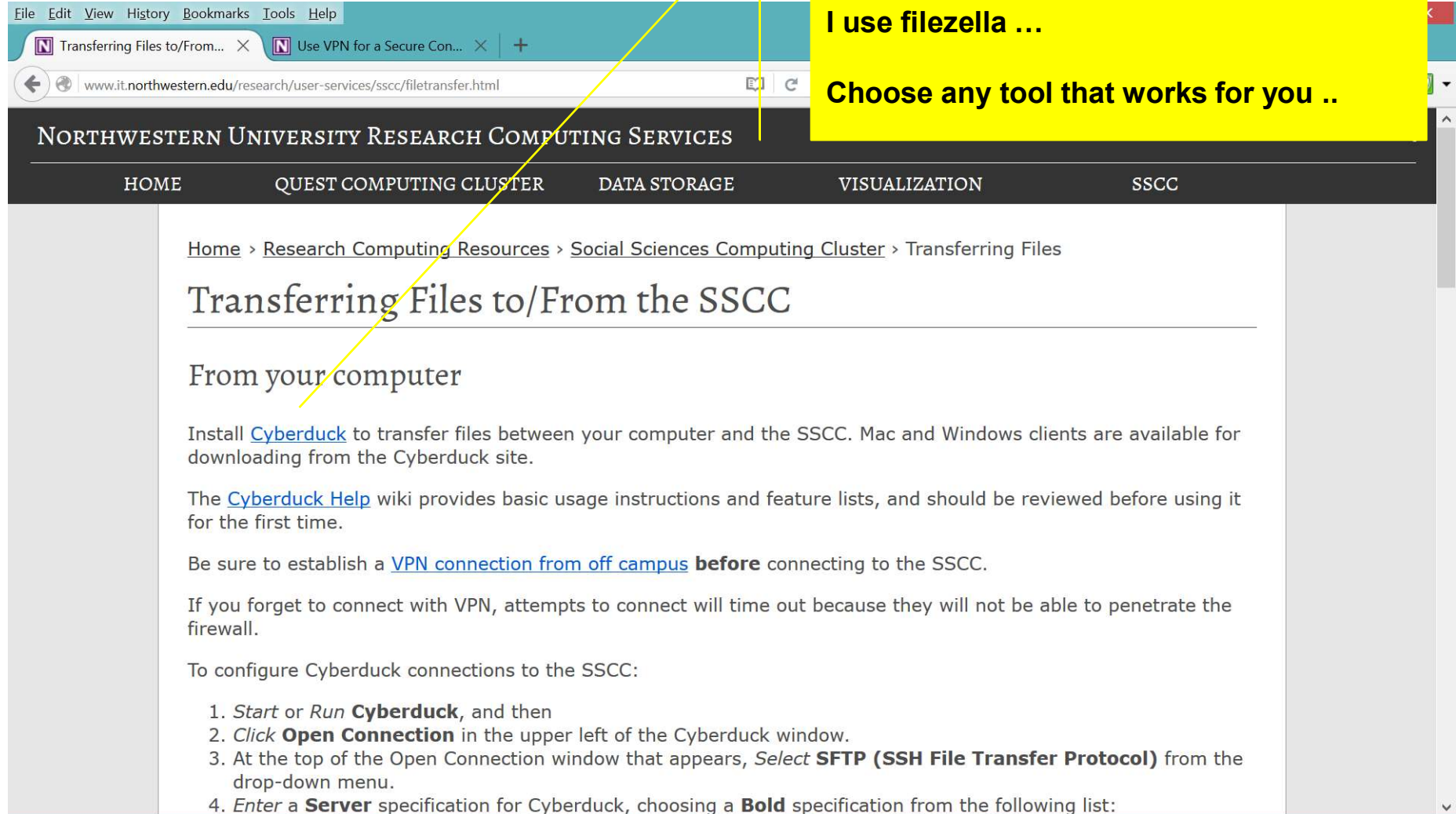


# PuTTY– Connecting to SSCC





# Bookmark this URL – Transferring Files



File Edit View History Bookmarks Tools Help

Transferring Files to/From... Use VPN for a Secure Con...

www.it.northwestern.edu/research/user-services/sscc/filetransfer.html

NORTHWESTERN UNIVERSITY RESEARCH COMPUTING SERVICES

HOME QUEST COMPUTING CLUSTER DATA STORAGE VISUALIZATION SSCC

Home > [Research Computing Resources](#) > [Social Sciences Computing Cluster](#) > Transferring Files

## Transferring Files to/From the SSCC

### From your computer

Install [Cyberduck](#) to transfer files between your computer and the SSCC. Mac and Windows clients are available for downloading from the Cyberduck site.

The [Cyberduck Help](#) wiki provides basic usage instructions and feature lists, and should be reviewed before using it for the first time.

Be sure to establish a [VPN connection from off campus](#) **before** connecting to the SSCC.

If you forget to connect with VPN, attempts to connect will time out because they will not be able to penetrate the firewall.

To configure Cyberduck connections to the SSCC:

1. **Start or Run Cyberduck**, and then
2. Click **Open Connection** in the upper left of the Cyberduck window.
3. At the top of the Open Connection window that appears, **Select SFTP (SSH File Transfer Protocol)** from the drop-down menu.
4. Enter a **Server** specification for Cyberduck, choosing a **Bold** specification from the following list:

**Cyberduck to transfer files**

**I use filezilla ...**

**Choose any tool that works for you ..**

# Bookmark this URL - Filezela

The screenshot shows a web browser window with the FileZilla Client Tutorial page. The browser's address bar displays the URL `https://wiki.filezilla-project.org/FileZilla_Client_Tutorial_(en)`. The page features a red FileZilla logo on the left, a navigation menu with links like 'Main Page', 'Community portal', and 'Recent changes', and a search bar. The main content area is titled 'FileZilla Client Tutorial (en)' and includes a table of contents on the right. The tutorial text explains how to connect to an FTP server, providing example login data: Hostname: example.org, Username: john, Password: 7PjU#.J3. It also shows a quickconnect bar with fields for Host, Username, Password, and Port, and a 'Quickconnect' button.

File Edit View History Bookmarks Tools Help

Research Computing Cons... FileZilla Client Tutorial (en)... Use VPN for a Secure Conn... +

https://wiki.filezilla-project.org/FileZilla\_Client\_Tutorial\_(en) Search

create account log in

**FileZilla Client Tutorial (en)**

This tutorial in other languages: [English \(English\)](#), [Deutsch \(German\)](#), [Español \(Spanish\)](#), [Français \(French\)](#)

*Note to editors: Please read [this note on the tutorial scope](#) before making additions.*

**Introduction**

Welcome to the FileZilla Client tutorial. In this tutorial, you will learn how to

- Connect to an FTP server,
- Download and upload files,
- Use the site manager.

If you already know how to use an FTP client, you may want to read the more advanced [usage instructions](#) instead.

We assume that you already installed and started FileZilla Client ([installation instructions](#)).

**Connecting to a server**

The first thing to do is connecting to a server.

This is our (fictional) login data - please use your own data instead if you want to actively follow the tutorial.

```
Hostname: example.org
Username: john
Password: 7PjU#.J3
```

We will use the quickconnect bar for establishing the connection:

Host:  Username:  Password:  Port:

Enter the hostname into the quickconnect bar's *Host*: field, the username into the *Username*: field as well as the password into the *Password*: field. You may leave the *Port*: field empty unless your login information specifies a certain port to use. Now click on *Quickconnect*.

**Note:** If your login information specifies a protocol like *SFTP* or *FTPS*, enter the hostname as follows: [sftp://hostname](#) or [ftps://hostname](#) respectively.

FileZilla will now try to connect to the server. If all works well, you will notice that the right "column" switched from *Not connected to any server* to displaying a list of files and directories.

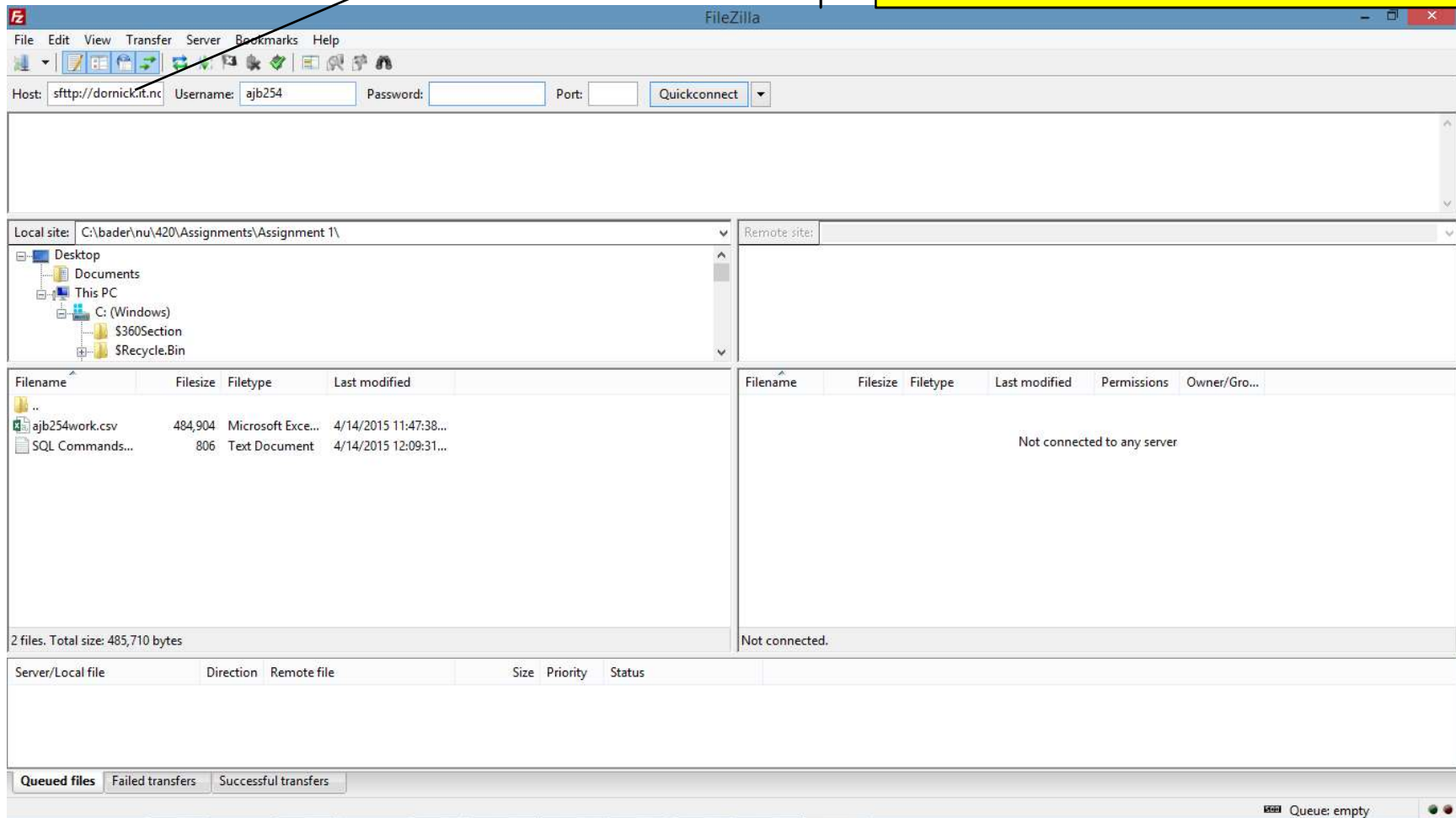
**Contents [hide]**

- 1 Introduction
- 2 Connecting to a server
- 3 Navigating and window layout
- 4 Transferring files
  - 4.1 Uploading
  - 4.2 Downloading
- 5 Using the site manager
- 6 Conclusion
- 7 Further information

# Filezela

**sftp://dornick.it.northwestern.edu**

**Enter sftp protocol, host, login, and password**



# Filezela

I'm connected now to :

sftp://dornick.it.northwestern.edu

The screenshot shows the FileZilla SFTP client interface. The title bar indicates the connection to `sftp://ajb254@dornick.it.northwestern.edu - FileZilla`. The top menu bar includes File, Edit, View, Transfer, Server, Bookmarks, and Help. Below the menu is a toolbar with various icons. The connection fields show Host: `sftp://dornick.it.no`, Username: `ajb254`, Password: masked with dots, and Port: empty. A 'Quickconnect' button is visible. The status bar at the bottom shows 'Queue: empty'.

**Status Messages:**

```
Status: Retrieving directory listing...
Status: Listing directory /sscc/home/a/ajb254
Status: Directory listing of "/sscc/home/a/ajb254" successful
Status: Retrieving directory listing of "/sscc/home/a/ajb254/ia1"...
Status: Listing directory /sscc/home/a/ajb254/ia1
Status: Directory listing of "/sscc/home/a/ajb254/ia1" successful
```

**Local site:** `C:\bader\nu\420\Assignments\Assignment 1\`

Filename	Filesize	Filetype	Last modified
..			
ajb254work.csv	484,904	Microsoft Excel...	4/14/2015 11:47:38...
SQL Commands...	806	Text Document	4/14/2015 12:09:31...

2 files. Total size: 485,710 bytes

**Remote site:** `/sscc/home/a/ajb254/ia1`

Filename	Filesize	Filetype	Last modified	Permissions	Owner/Gro...
..					
ajb254w...	484,904	Microsoft ...	4/14/2015 11:1...	-rw-----	ajb254 users

1 file. Total size: 484,904 bytes

**Transfer Queue:**

Server/Local file	Direction	Remote file	Size	Priority	Status
-------------------	-----------	-------------	------	----------	--------

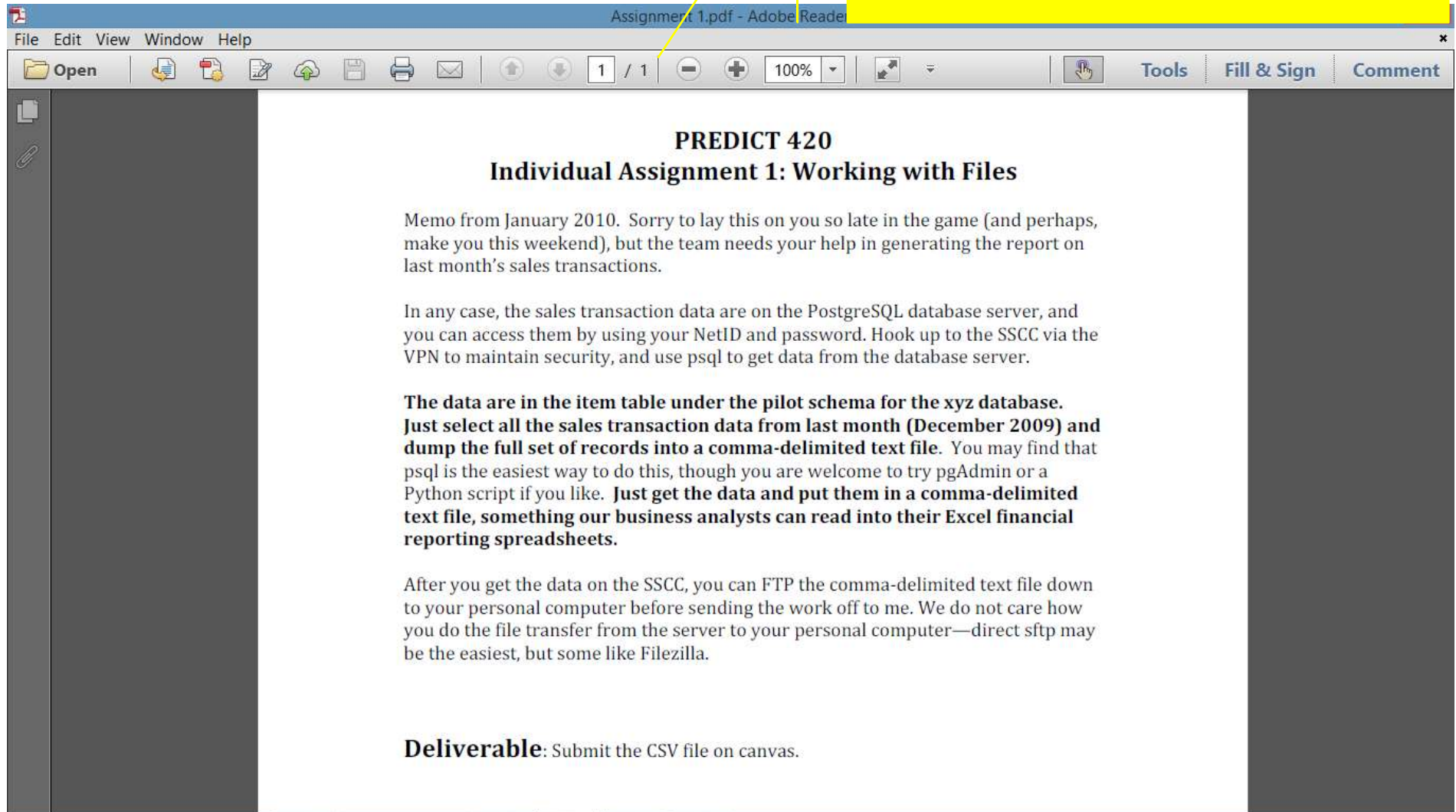
Queued files | Failed transfers | Successful transfers

---

# Assignment #1

# Assignment #1

## Assignment 1 Instructions & Deliverable



The screenshot shows the Adobe Reader interface with a document titled "Assignment 1.pdf". The document content is as follows:

**PREDICT 420**  
**Individual Assignment 1: Working with Files**

Memo from January 2010. Sorry to lay this on you so late in the game (and perhaps, make you this weekend), but the team needs your help in generating the report on last month's sales transactions.

In any case, the sales transaction data are on the PostgreSQL database server, and you can access them by using your NetID and password. Hook up to the SSCC via the VPN to maintain security, and use psql to get data from the database server.

**The data are in the item table under the pilot schema for the xyz database. Just select all the sales transaction data from last month (December 2009) and dump the full set of records into a comma-delimited text file.** You may find that psql is the easiest way to do this, though you are welcome to try pgAdmin or a Python script if you like. **Just get the data and put them in a comma-delimited text file, something our business analysts can read into their Excel financial reporting spreadsheets.**

After you get the data on the SSCC, you can FTP the comma-delimited text file down to your personal computer before sending the work off to me. We do not care how you do the file transfer from the server to your personal computer—direct sftp may be the easiest, but some like Filezilla.

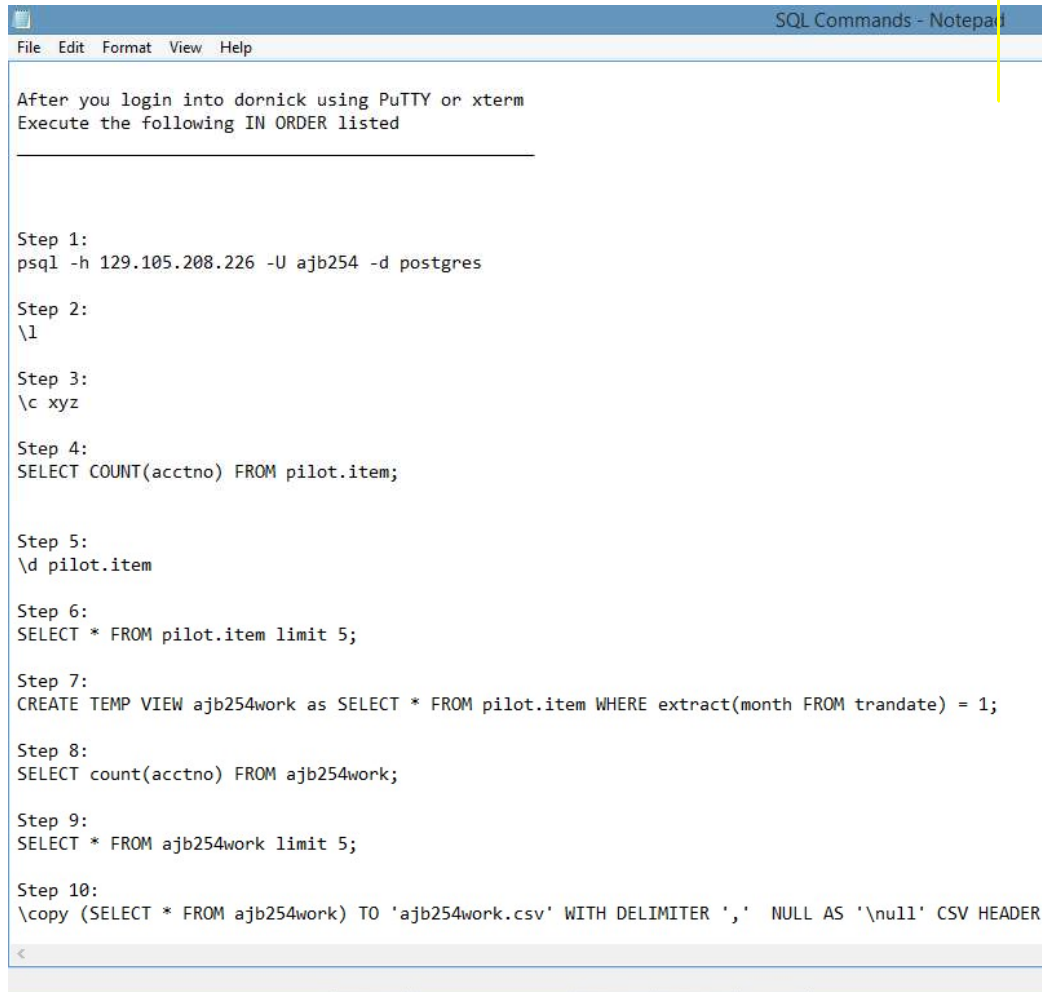
**Deliverable:** Submit the CSV file on canvas.



# Assignment #1

This file has the set of steps that you could use to type these commands

Step-by-step



```
File Edit Format View Help
SQL Commands - Notepad

After you login into dornick using PuTTY or xterm
Execute the following IN ORDER listed

Step 1:
psql -h 129.105.208.226 -U ajb254 -d postgres

Step 2:
\l

Step 3:
\c xyz

Step 4:
SELECT COUNT(acctno) FROM pilot.item;

Step 5:
\d pilot.item

Step 6:
SELECT * FROM pilot.item limit 5;

Step 7:
CREATE TEMP VIEW ajb254work as SELECT * FROM pilot.item WHERE extract(month FROM trandate) = 1;

Step 8:
SELECT count(acctno) FROM ajb254work;

Step 9:
SELECT * FROM ajb254work limit 5;

Step 10:
\copy (SELECT * FROM ajb254work) TO 'ajb254work.csv' WITH DELIMITER ',' NULL AS '\null' CSV HEADER
```





Ln 1, Col 1

# Assignment #1

---



**Set of files describe the XYZ database**

-  XYZ\_Data Dictionaries
-  XYZ\_Data Dictionaries\_Customer\_Table
-  XYZ\_Data Dictionaries\_Item\_Table
-  XYZ\_Data Dictionaries\_Mail\_Table



# Assignment #1 - Deliverable

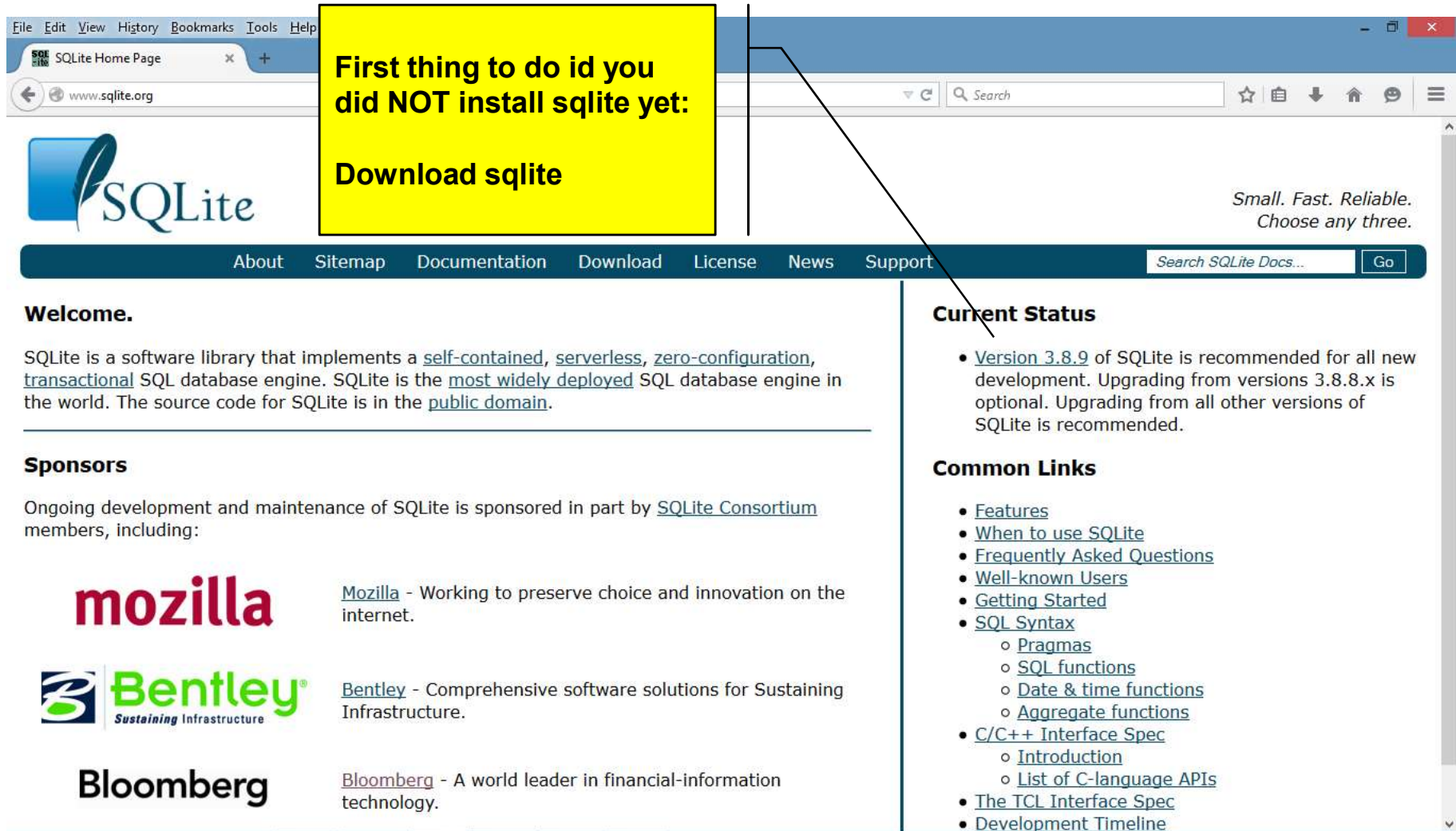
---

- ❑ **After you execute all SQL commands and ftp the file from dornick server**
- ❑ **Submit the CSV file on canvas.**

---

# Exercises #3

# Exercise #3 – Download SQLite



The screenshot shows the SQLite Home Page in a web browser. A yellow callout box with the text "First thing to do if you did NOT install sqlite yet: Download sqlite" is overlaid on the page. A line points from the callout box to the "Download" link in the navigation bar.

**SQLite Home Page**

www.sqlite.org

SQLite

Small. Fast. Reliable.  
Choose any three.

About Sitemap Documentation Download License News Support

Search SQLite Docs... Go

### Welcome.

SQLite is a software library that implements a [self-contained](#), [serverless](#), [zero-configuration](#), [transactional](#) SQL database engine. SQLite is the [most widely deployed](#) SQL database engine in the world. The source code for SQLite is in the [public domain](#).

### Sponsors

Ongoing development and maintenance of SQLite is sponsored in part by [SQLite Consortium](#) members, including:

**mozilla** [Mozilla](#) - Working to preserve choice and innovation on the internet.

**Bentley** [Bentley](#) - Comprehensive software solutions for Sustaining Infrastructure.

**Bloomberg** [Bloomberg](#) - A world leader in financial-information technology.

### Current Status

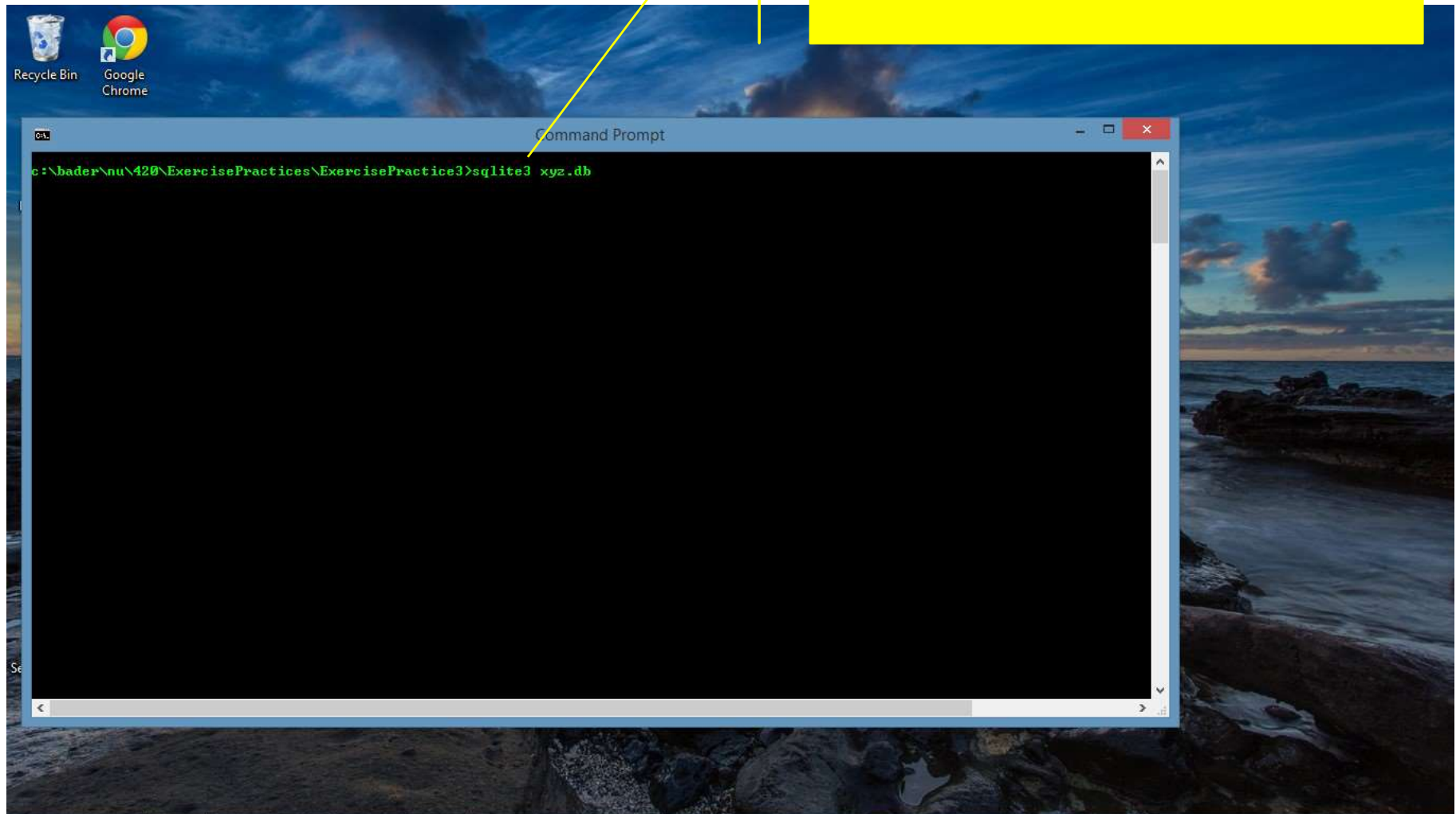
- [Version 3.8.9](#) of SQLite is recommended for all new development. Upgrading from versions 3.8.8.x is optional. Upgrading from all other versions of SQLite is recommended.

### Common Links

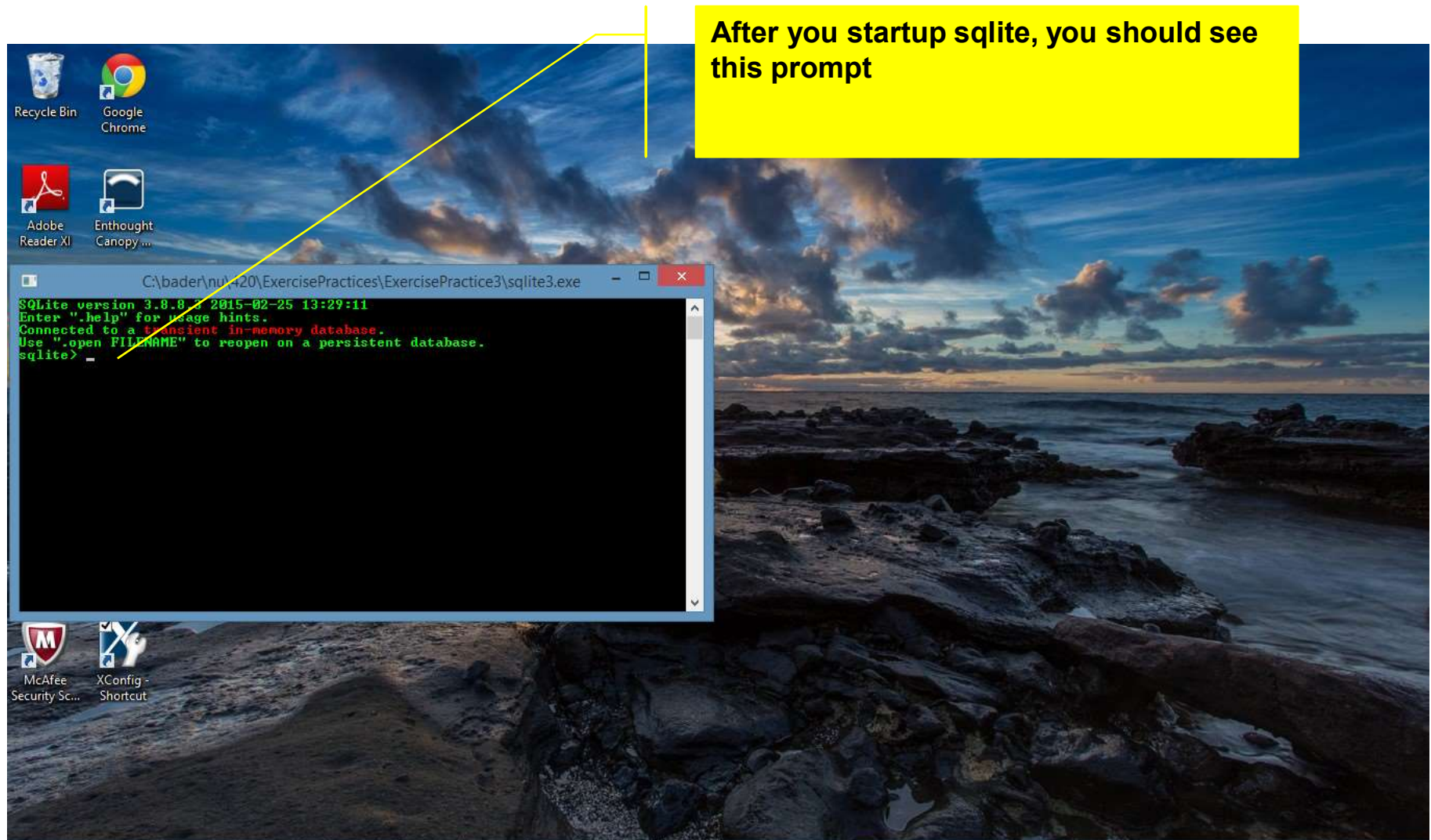
- [Features](#)
- [When to use SQLite](#)
- [Frequently Asked Questions](#)
- [Well-known Users](#)
- [Getting Started](#)
- [SQL Syntax](#)
  - [Pragmas](#)
  - [SQL functions](#)
  - [Date & time functions](#)
  - [Aggregate functions](#)
- [C/C++ Interface Spec](#)
  - [Introduction](#)
  - [List of C-language APIs](#)
- [The TCL Interface Spec](#)
- [Development Timeline](#)

# Exercise #3 – Run SQLite

From command prompt type the following to startup sqlite



# Exercise #3 – Run SQLite



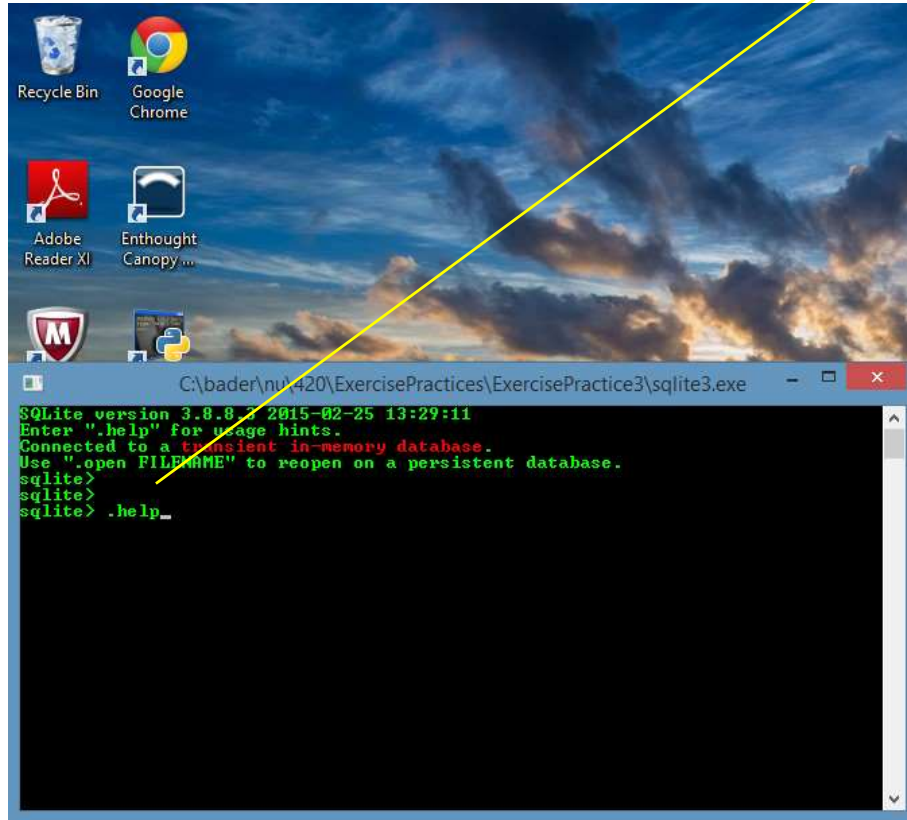


# Exercise #3 – Run SQLite

Type `.help`

(it is DOT help)

Hit enter to see list of commands



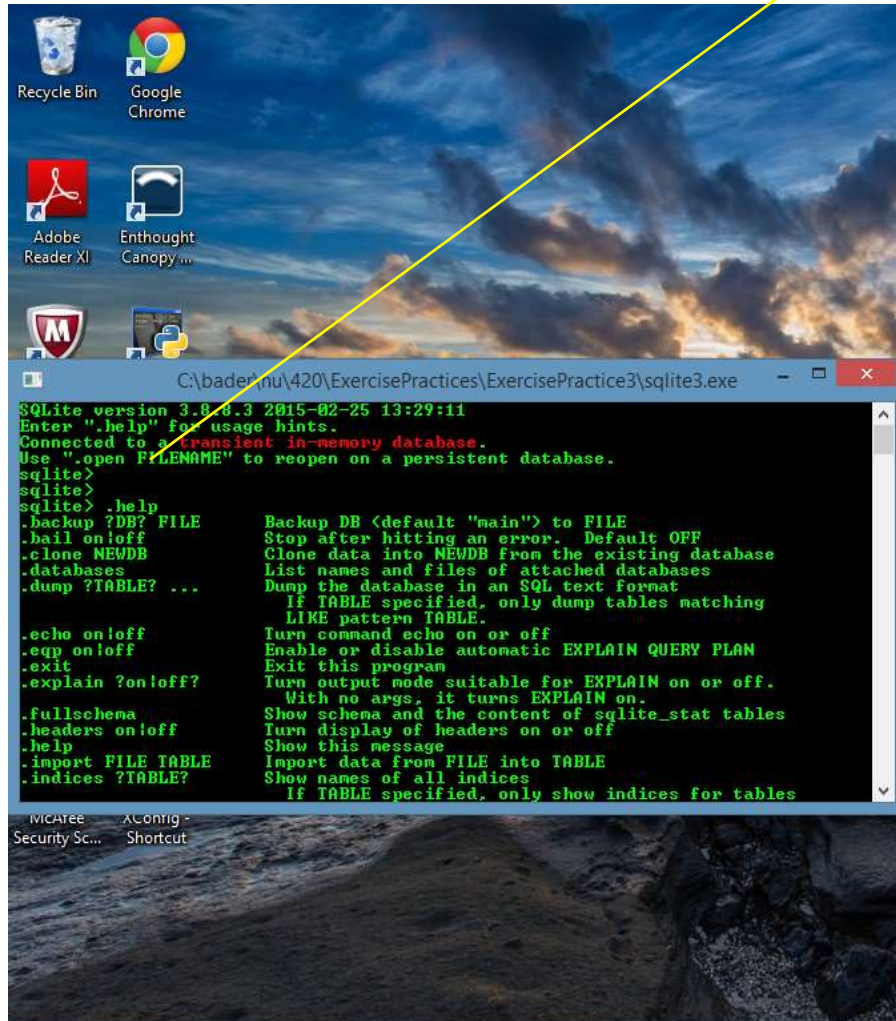
The screenshot shows a Windows desktop with a scenic background of a rocky coastline at sunset. In the top-left corner, there are icons for the Recycle Bin, Google Chrome, Adobe Reader XI, and Enthought Canopy. A command prompt window titled "C:\bader\nu\420\ExercisePractices\ExercisePractice3\sqlite3.exe" is open in the bottom-left. The window displays the following text:

```
SQLite version 3.8.7.2 2015-02-25 13:29:11
Enter ".help" for usage hints.
Connected to a transient in-memory database.
Use ".open FILENAME" to reopen on a persistent database.
sqlite>
sqlite>
sqlite> .help_
```

MCaree - AConrig -  
Security Sc... Shortcut

# Exercise #3 – Run SQLite

Now you should see list of commands



The screenshot shows a Windows desktop with a scenic background of a rocky coastline at sunset. In the top-left corner, there are icons for the Recycle Bin, Google Chrome, Adobe Reader XI, and Enthought Canopy. A command prompt window titled "C:\bader\hu\420\ExercisePractices\ExercisePractice3\sqlite3.exe" is open, displaying the SQLite version 3.8.8.3 and the time 2015-02-25 13:29:11. The prompt shows the user has entered ".help", and the command prompt displays a list of available commands and their descriptions. A yellow line points from the text box above to the ".help" command in the command prompt.

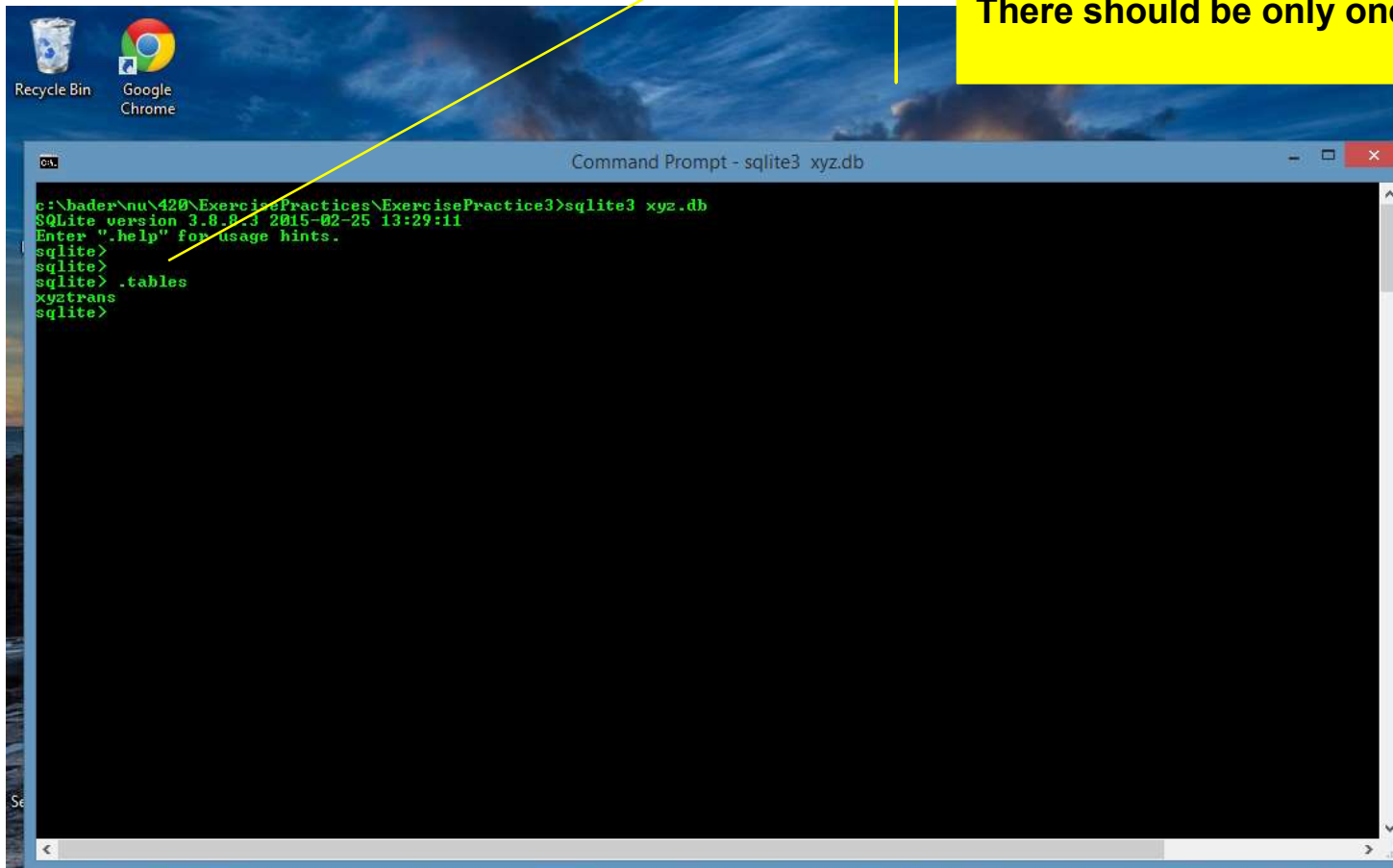
```
SQLite version 3.8.8.3 2015-02-25 13:29:11
Enter ".help" for usage hints.
Connected to a transient in-memory database.
Use ".open FILENAME" to reopen on a persistent database.
sqlite>
sqlite>
sqlite>.help
.backup ?DB? FILE      Backup DB (default "main") to FILE
.bail on|off           Stop after hitting an error. Default OFF
.clone NEWDB            Clone data into NEWDB from the existing database
.databases              List names and files of attached databases
.dump ?TABLE? ...      Dump the database in an SQL text format
                        If TABLE specified, only dump tables matching
                        LIKE pattern TABLE.
.echo on|off           Turn command echo on or off
.explain on|off         Enable or disable automatic EXPLAIN QUERY PLAN
.exit                  Exit this program
.explain ?on|off?       Turn output mode suitable for EXPLAIN on or off.
                        With no args, it turns EXPLAIN on.
.fullschema             Show schema and the content of sqlite_stat tables
.headers on|off         Turn display of headers on or off
.help                  Show this message
.import FILE TABLE     Import data from FILE into TABLE
.indices ?TABLE?        Show names of all indices
                        If TABLE specified, only show indices for tables
```



# Exercise #3 – Run SQLite

Type `.tables`

There should be only one table: `xyztrans`



```
c:\bader\nu\420\ExercisePractices\ExercisePractice3>sqlite3 xyz.db
SQLite version 3.8.8.3 2015-02-25 13:29:11
Enter ".help" for usage hints.
sqlite>
sqlite>
sqlite> .tables
xyztrans
sqlite>
```



# Exercise #3 – Cinstall SQLAlchemy Package

Make sure you install sqlalchemy package before you run the script.

The screenshot shows a Jupyter Notebook titled 'exercise3' running on a local host. The notebook contains text explaining that SQLAlchemy is needed for pandas and that it should be installed. A code cell shows the command `import sqlalchemy`, which has failed, resulting in a `ImportError: No module named sqlalchemy`. A purple arrow points from the instructional text box to the failed import line. Below the error, the notebook continues with instructions on how to use SQLAlchemy's `create_engine` function to connect to a SQLite database.

File Edit View History Bookmarks Tools Help

localhost:8888/notebooks/bader/nu/420/spring 2016/sync sessions/sync session 3/exercisepractice3/exer

jupyter exercise3 Last Checkpoint: 04/10/2015 (unsaved changes)

Python 2

File Edit View Insert Cell Kernel Help

Code Cell Toolbar: None

There are a couple of different ways to read and write data to RDBs using Python, but the most flexible and easiest may be by using what's in pandas. pandas will make use of the SQLAlchemy package, which is available for installation within Canopy. (Did you install it in Session 1?) SQLAlchemy provides a consistent interface with different RDBs, SQLite being one of them.

Let's get SQLAlchemy into our IPython session:

```
In [15]: import sqlalchemy
```

-----

```
ImportError                                Traceback (most recent call last)
<ipython-input-15-9b4dcd646713> in <module>()
----> 1 import sqlalchemy

ImportError: No module named sqlalchemy
```

Now if you do the `sqlalchemy.<tab>` trick from the command prompt, you'll be able to see SQLAlchemy's various (and many) attributes and functions.

To simplify things, let's get a function out of SQLAlchemy that we'll use to define the SQLite3 db we'll be working with:

```
In [ ]: from sqlalchemy import create_engine
```

Now let's specify the xyz db as the SQLite3 RDB we want to work with:

```
In [ ]: engine=create_engine('sqlite:///xyz.db')
```

This assumes that you have xyz.db in your current working directory. There are different valid syntaxes, e.g.

# Exercise #3 – Install SQLAlchemy Package

Make sure you install sqlalchemy package before you run the script.

The screenshot shows the Canopy Package Manager interface. The title bar reads "Package Manager - Canopy". The menu bar includes "File", "Edit", "Tools", "Window", and "Help". A search bar at the top left contains the text "sqlal". On the right side of the header, there is a "CANOPY" logo, the name "Atef Bader", and a link "Need help? Ask experts". The main area displays a list of available packages. The top package, "SQLAlchemy 1.0.12-1", is highlighted in blue. Its details show it is an "SQL toolkit and Object Relational Mapper" with version "1.0.12" and size "3.07 MB". Below this, a list of other packages is shown, including "zope.sqlalchemy 0.7.6-6", "clue\_sqlaloader 0.1.1-1", "cromlech.sqlalchemy 0.2a1-1", "khufu\_sqlalchemy 1.0a1-1", "pas.plugins.sqlalchemy 0.3-1", "Products.SQLAlchemyDA 0.5.1-1", "Products.SQLAlchemyPAS 1.0-1", "sqlalchemy\_admin 0.1-1", "sqlalchemy\_schemadisplay 1.1-1", "SQLAlchemyBWC 0.2.5-1", "SQLAlchemyManager 0.1.0-1", "tiddlywebplugins.sqlalchemy 0.9.18-1", "tiddlywebplugins.sqlalchemy2 2.0.12-1", "transmogrify.sqlalchemy 1.0.1-1", and "z3c.sqlalchemy 1.4.0-1". Each package has a small icon and a "Pyth" label. At the bottom left, it says "14397 packages available. 16 matches". At the bottom right, there is a button that says "Install all Canopy packages".

File Edit Tools Window Help

sqlal

SQLAlchemy 1.0.12-1  
SQL toolkit and Object Relational Mapper  
Version: 1.0.12  
Size: 3.07 MB

zope.sqlalchemy 0.7.6-6  
clue\_sqlaloader 0.1.1-1  
cromlech.sqlalchemy 0.2a1-1  
khufu\_sqlalchemy 1.0a1-1  
pas.plugins.sqlalchemy 0.3-1  
Products.SQLAlchemyDA 0.5.1-1  
Products.SQLAlchemyPAS 1.0-1  
sqlalchemy\_admin 0.1-1  
sqlalchemy\_schemadisplay 1.1-1  
SQLAlchemyBWC 0.2.5-1  
SQLAlchemyManager 0.1.0-1  
tiddlywebplugins.sqlalchemy 0.9.18-1  
tiddlywebplugins.sqlalchemy2 2.0.12-1  
transmogrify.sqlalchemy 1.0.1-1  
z3c.sqlalchemy 1.4.0-1

14397 packages available. 16 matches

Install all Canopy packages

# Exercise #3 – Ipython Notebook Script

Run Ipython Notebook Script

The screenshot shows the Canopy IDE interface. The main editor window displays a notebook titled "Week #3 - Python Practice". The notebook content includes a text block explaining the task and a code cell with Python imports. The bottom panel shows the Python interpreter output.

**Week #3 - Python Practice**

Where we left off in the last session is with some questions about XYZ's customer data. We pickled the data we had input from a csv file. (We did, right?) Let's get started again in Canopy and load our pickled data. The first thing to do is to import pandas, numpy, DataFrame from pandas, and cPickle into Python. Remember the nicknames we gave some of them when we imported them? I'm talking about pd, np, and pickle. They're optional, but they're convenient and also used by many Python programmers.

Anyway, assuming that you stored your XYZ customer data in a pickle file called xyzcust10.p, you can unpickle it into Python like:

```
In [180]: import os
import cPickle as pickle

import pandas as pd # panda's nickname is pd
import numpy as np # numpy as np
from pandas import DataFrame, Series # for convenience
```

Python

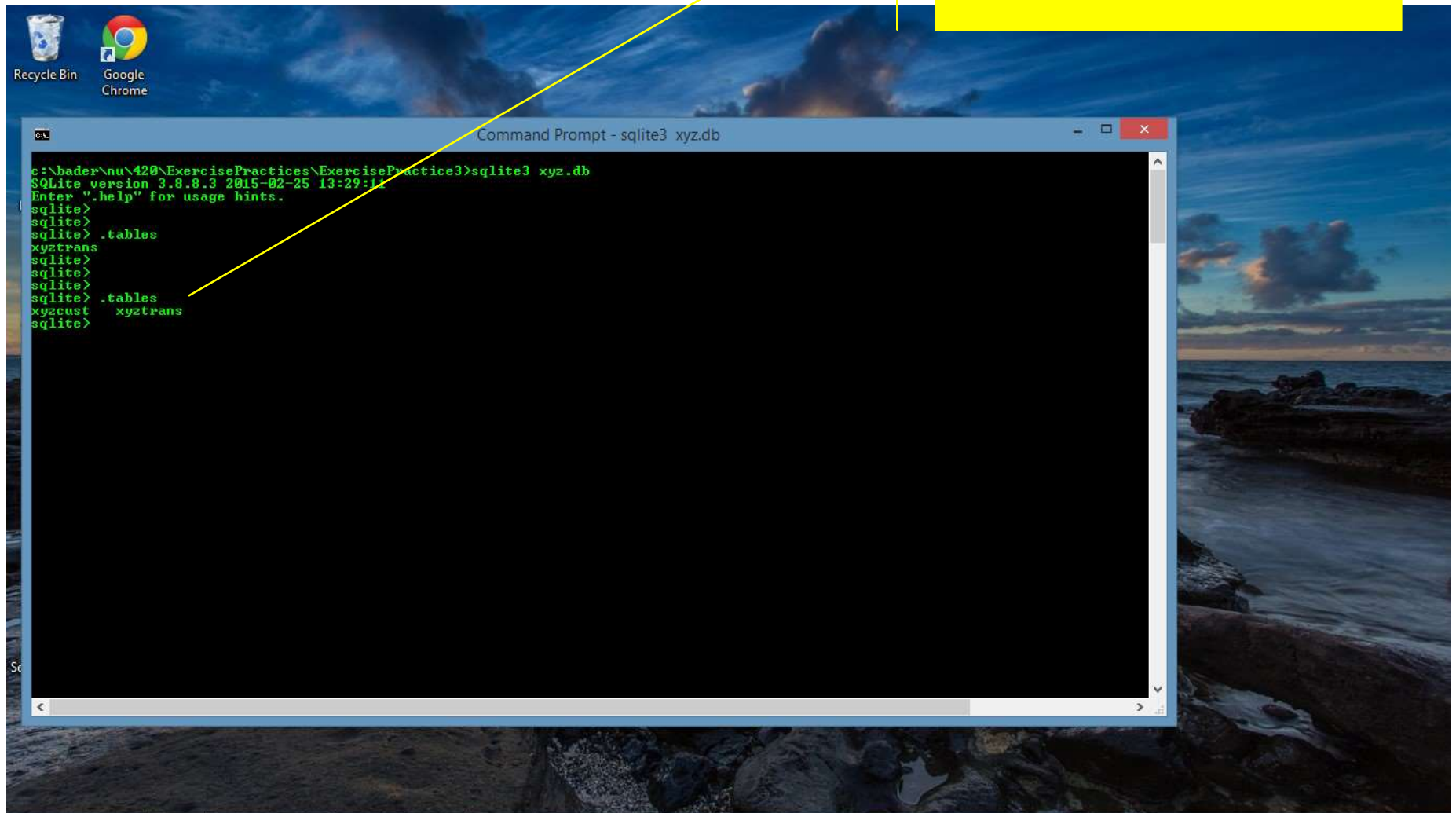
Welcome to Canopy's interactive data-analysis environment!  
with pylab-backend set to: qt  
Type '?' for more information.

In [1]:

C:\bader\nu420\ExercisePractices\ExercisePractice3\Exercise3.ipynb

# Exercise #3 – Ipython Notebook Script

Type .tables  
You should see TWO tables  
xyzcust and xyztrans



```
c:\bader\nu\420\ExercisePractices\ExercisePractice3>sqlite3 xyz.db
SQLite version 3.8.8.3 2015-02-25 13:29:14
Enter ".help" for usage hints.
sqlite>
sqlite>
sqlite> .tables
xyztrans
sqlite>
sqlite>
sqlite>
sqlite> .tables
xyzcust  xyztrans
sqlite>
```

# Exercise #3 – Ipython Notebook Script

Run Ipython Notebook Script  
AGAIN

Did you get the following error? Why?

The screenshot shows the Canopy IDE interface. The main editor window displays a Python script with the following code:

```
C:\Users\user1\AppData\Local\Enthought\Canopy\User\lib\site-packages\pandas\io\sql.py in to_sql(self, frame, name, if_exists, index, index_label, schema, chunksize, dtype)
1174         if_exists=if_exists, index_label=index_label,
1175         schema=schema, dtype=dtype)
-> 1176     table.create()
1177     table.insert(chunksize)
1178     # check for potentially case sensitivity issues (GH7815)

C:\Users\user1\AppData\Local\Enthought\Canopy\User\lib\site-packages\pandas\io\sql.py in create(self)
637     if self.exists():
638         if self.if_exists == 'fail':
--> 639             raise ValueError("Table '%s' already exists." % self.name)
640         elif self.if_exists == 'replace':
641             self.pd_sql.drop_table(self.name, self.schema)

ValueError: Table 'xyzcust' already exists.
```

Below the code, there is a text input field with the text "Did it create the table in xyz.db? Check:". Below that, there is a code input field with the text "In [ ]: pd.read\_sql\_table('xyzcust', engine).columns". Below the code input field, there is a text input field with the text "should produce the columns of the DataFrame you wrote to the db. Remember that 'engine' refers to the SQLite3 DB by way of defining the connection using SQLAlchemy's create\_engine method."

The bottom panel shows the Python terminal with the following text:

```
Python
Welcome to Canopy's interactive data-analysis environment!
with pylab-backend set to: qt
Type '?' for more information.

In [1]:
```

The taskbar at the bottom shows the Windows Start button, several application icons, and the system tray with the date and time "8:06 PM" and "C:\bader\nu\420\ExercisePractices\ExercisePractices\Exer4\14\2015.b".



# Exercise #3 – Ipython Notebook Script

drop xyzcust table

And then do .tables

Only one table left



```
c:\bader\nu\420\ExercisePractices\ExercisePractice3\sqlite3 xyz.db
SQLite version 3.8.8.3 2015-02-25 13:29:11
Enter ".help" for usage hints.
sqlite>
sqlite>
sqlite> .tables
xyztrans
sqlite>
sqlite>
sqlite>
sqlite> .tables
xyzcust  xyztrans
sqlite>
sqlite> drop table xyzcust;
sqlite>
sqlite> .tables
xyztrans
sqlite>
```

# Exercise #3 – Ipython Notebook Script

This URL has a set of Simple SQL examples for SQLite

The screenshot shows a web browser window with the URL `www.tutorialspoint.com/sqlite/sqlite_drop_table.htm`. The page title is "SQLite - DROP Table". The website header includes the "tutorialspoint" logo and the tagline "SIMPLY EASY LEARNING". A navigation menu lists various topics: Home, Programming, Java, Web, Databases, Academic, Management, Quality, Telecom, and More... There is also a search bar and links to REFERENCES, FORUM, ABOUT, and CONTACT.

The main content area is titled "SQLite - DROP Table" and includes an advertisement for internet service. Below the ad, there are links for "Previous Page" and "Next Page". The text explains that the SQLite **DROP TABLE** statement is used to remove a table definition and all associated data, indexes, triggers, constraints, and permission specifications for that table. It includes a warning: "You have to be careful while using this command because once a table is deleted then all the information available in the table would also be lost forever."

**Syntax:**  
Basic syntax of DROP TABLE statement is as follows. You can optionally specify database name along with table name as follows:

```
DROP TABLE database_name.table_name;
```

**Example:**  
Let us first verify COMPANY table and then we would delete it from the database.

```
sqlite>.tables
COMPANY      test.COMPANY
```

This means COMPANY table is available in the database, so let us drop it as follows:

```
sqlite>DROP TABLE COMPANY;
sqlite>
```

The sidebar on the left contains a "Learn SQLite" section with a list of links: SQLite - Home, SQLite - Overview, SQLite - Installation, SQLite - Commands, SQLite - Syntax, SQLite - Data Types, SQLite - Create Database, SQLite - Attach Database, SQLite - Detach Database, SQLite - Create Table, **SQLite - Drop Table**, SQLite - Insert Query, SQLite - Select Query, SQLite - Operators, SQLite - Expressions, SQLite - Where Clause, and SQLite - AND & OR Clauses.

On the right side, there are social media links (Facebook, Twitter, Google+, LinkedIn, Email, RSS), links to "Modern Baby Names" and "Online Photo Editing", and an advertisement for "ObjectRocket by Rackspace" with the text "Automated: sharding, scaling, backups, compaction,".

# Exercise #3 – Ipython Notebook Script

Run Ipython Notebook Script  
After you drop the xyzcust  
table from sqlite/xyz.db

Did you get an error now?

You should get NO error now

The screenshot shows the Canopy IDE interface. On the left is a File Browser showing 'user1' and 'Recent Files'. The main editor window is titled 'Exercise3.ipynb' and contains the following text:

### Week #3 - Python Practice

Where we left off in the last session is with some questions about XYZ's customer data. We pickled the data we had input from a csv file. (We did, right?) Let's get started again in Canopy and load our pickled data. The first thing to do is to import pandas, numpy, DataFrame from pandas, and cPickle into Python. Remember the nicknames we gave some of them when we imported them? I'm talking about pd, np, and pickle. They're optional, but they're convenient and also used by many Python programmers.

Anyway, assuming that you stored your XYZ customer data in a pickle file called xyzcust10.p, you can unpickle it into Python like:

```
In [180]: import os  
  
import cPickle as pickle  
  
import pandas as pd # panda's nickname is pd  
  
import numpy as np # numpy as np  
  
from pandas import DataFrame, Series # for convenience
```

The bottom panel shows the Python output:

```
Python  
Welcome to Canopy's interactive data-analysis environment!  
with pylab-backend set to: qt  
Type '?' for more information.  
  
In [1]:
```

C:\bader\nu\420\ExercisePractices\ExercisePractice3\Exercise3.ipynb



# Exercises #3 - Deliverable

---

- A single PDF document that has the following 6 screen-shots
  1. sqlite3 command-window shows only one table before you run the Ipython Notebook script
  2. First Run for the Ipython Notebook script with no error
  3. sqlite3 command-windows shows two tables after you run the Ipython Notebook script
  4. Second Run for the Ipython Notebook script that shows error
  5. sqlite3 command-window shows drop table for xyzcust
  6. Third Run for the Ipython Notebook script that shows NO error