MEDICAL MICROBIOLOGY AND INFECTIOUS DISEASES BIOINFORMATICS WORKSHOP

Presents

Introduction to the 2023 MMID Bioinformatics Workshop

INSTRUCTED BY

Grace E. Seo, MSc Student

INFORMATION FOR PARTICIPANTS

All workshops are being recorded and posted to the MMID Bioinformatics Workshop – YouTube

For live Q&A, go to slido.com and use participant code #4156734

2023 MMID Bioinformatics Workshop Schedule

DATE	INSTRUCTOR	TOPIC
March 2	Grace E. Seo	Introduction to the 2023 MMID Bioinformatics Workshop
March 9	Grace E. Seo	Introduction to conda and tool installation
March 16	Grace E. Seo	Introduction to genomics and viral data analysis
March 23	Jill Rumore	Bacterial Genomics
March 30	Jill Rumore	Reference Databases
April 6	Taylor Davedow	Beginner's Guide to Phylogenetic Trees
April 13	Taylor Davedow	Introduction to tree visualization and annotation using ggtree
April 20	-	Bfx workshop: Bring your own dataset!
April 27	-	Bfx workshop: Bring your own dataset!

April 20 and April 27 in-person sessions are open to the public (up to 100 people)!

Work with your colleagues/friends to analyze data together.

Objectives are hyperlinked for easy navigation within the document.

LEARNING OBJECTIVES

1. ***Explain bioinformatics***

2. ***Describe the application of bioinformatics in health science research***

3. ***Describe the benefits of using a command line interface (CLI) over a graphical user interface (GUI)***

4. ***Install BASH terminal and RStudio***

SET UP WI-FI (IN-PERSON PARTICIPANTS)

1. ***Connect to UofM-secure (if you are a student or staff)***
   - Use your @myumanitoba.ca or @umanitoba.ca login and password

2. ***Connect to UofM-guest***

To access uofm-guest Wi-Fi:

1. Ensure your wireless card is active and connected to the **uofm-guest** network.
2. Open your web browser (e.g. Google Chrome, Microsoft Edge, Firefox, etc.) and browse to any website. This should redirect you to the **Acceptable Use Agreement** page.
3. Review the Acceptable Use Agreement for the unsecured wireless.
4. Select **I Agree**.

LEARNING OBJECTIVES

1. *Explain bioinformatics*
2. *Describe the application of bioinformatics in health science research*
3. *Describe the benefits of using a command line interface (CLI) over a graphical user interface (GUI)*
4. *Install BASH terminal and RStudio*

BIOINFORMATICS

1. ***Bioinformatics (Biology + Informatics) is an application of computer technology to collect, store, analyze and interpret to understand biological data[1, 2].***

2. ***Bioinformatics is largely used in Genomics, Proteomics, Drug designing, etc. for big data analysis.***

3. ***Bioinformatics uses computer programming to improve software for more efficient and quicker data analysis with better accuracy.***

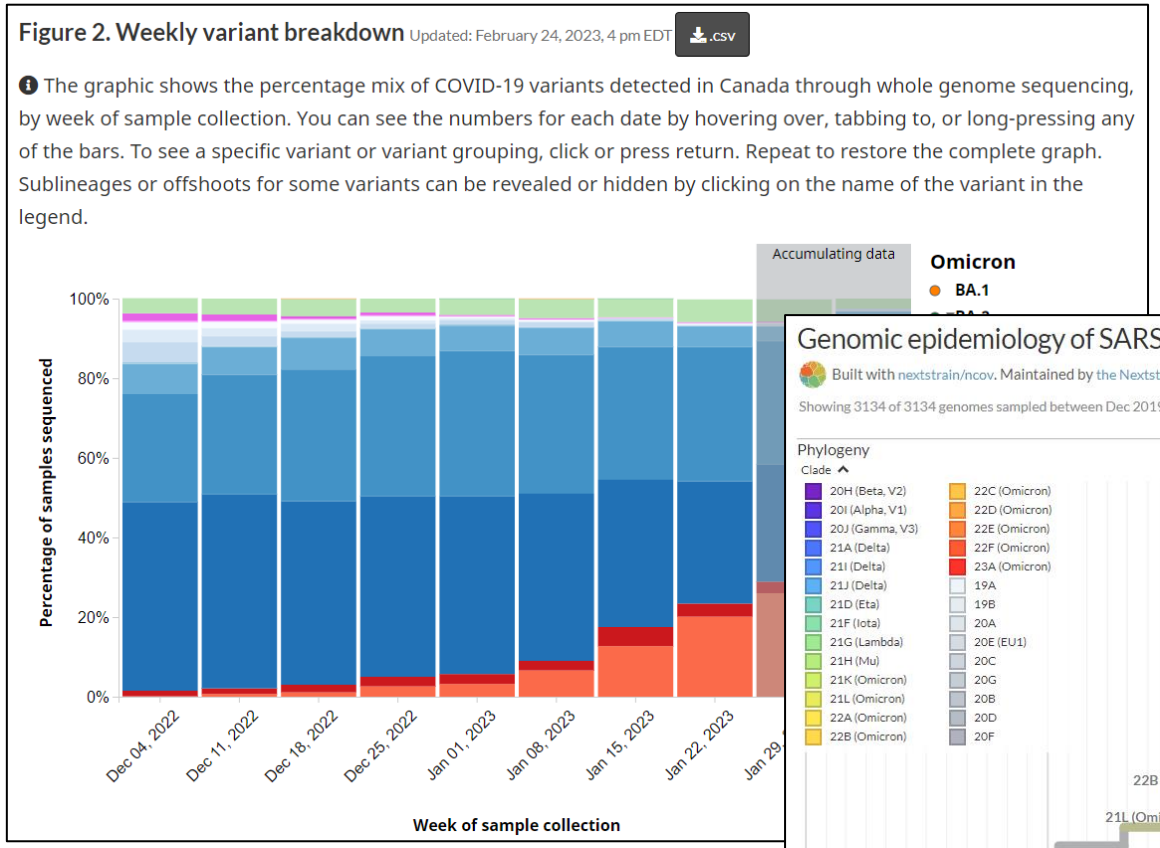
1. Adams, D. 2023. Bioinformatics. National Human Genome Research Institute. Available from https://www.genome.gov/genetics-glossary/Bioinformatics [accessed 2 March 2023].
2. Bayat, A. 2002. Bioinformatics. British Medical Journal. 324(7344): 1018-1022. doi: https://doi.org/10.1136%2Fbmj.324.7344.1018.

Click to go back to the beginning

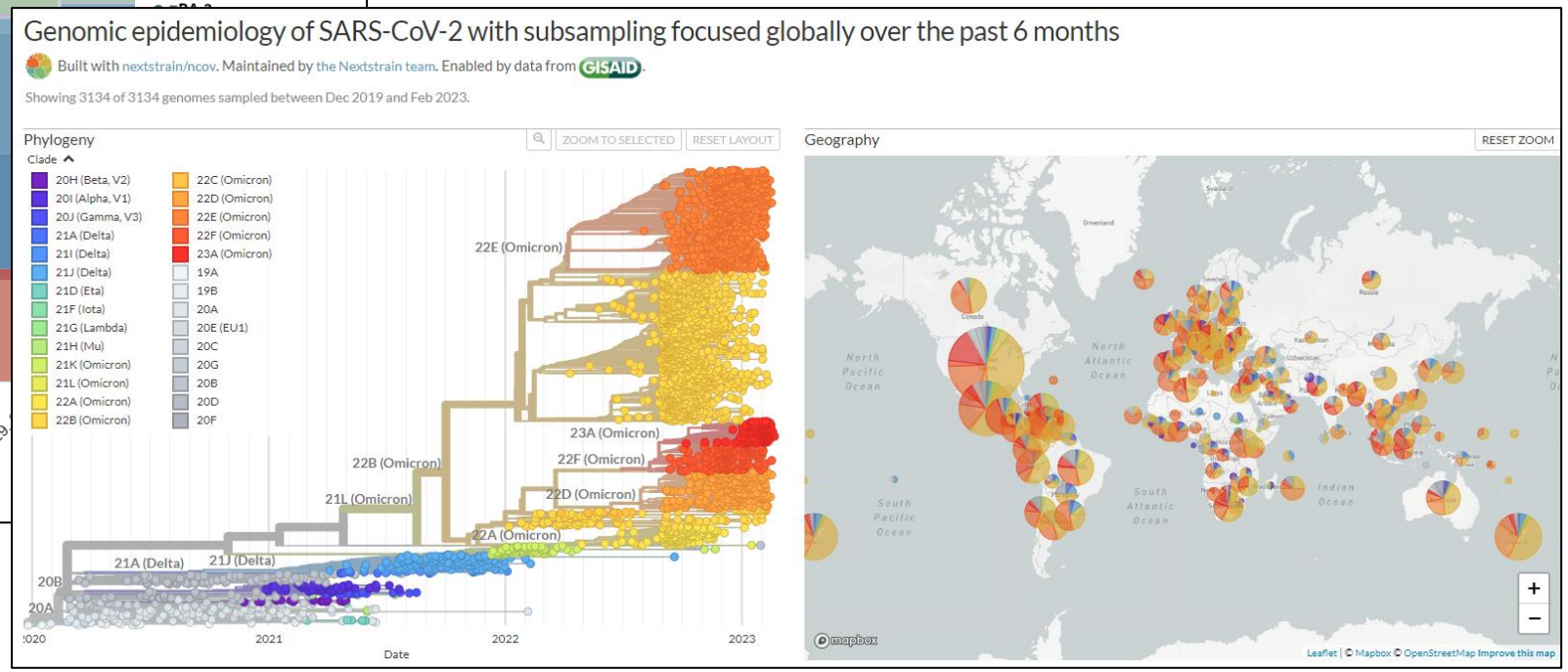
LEARNING OBJECTIVES

1. *Explain bioinformatics*
2. ***Describe the application of bioinformatics in health science research***
3. *Describe the benefits of using a command line interface (CLI) over a graphical user interface (GUI)*
4. *Install BASH terminal and RStudio*

SARS-CoV-2 Whole-genome sequencing



April 6 and 13th
- Taylor Davedow's ggtree workshop



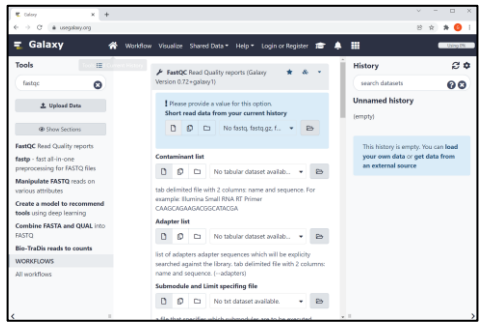
Click to go back to the beginning

LEARNING OBJECTIVES

1. *Explain bioinformatics*
2. *Describe the application of bioinformatics in health science research*
3. ***Describe the benefits of using a command line interface (CLI) over a graphical user interface (GUI)***
4. *Install BASH terminal and RStudio*

Graphical user interface (GUI) → Command line interface (CLI) → Computer = Result

GUI



https://usegalaxy.org/

CLI

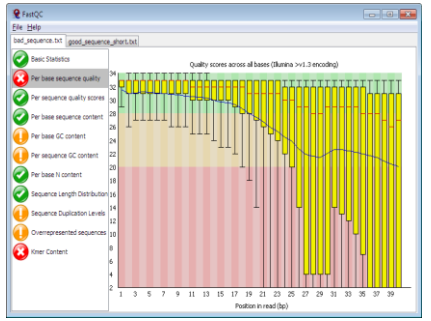
```
ubuntu@domU-12-31-39-02-88-C1: ~
File  Edit  View  Search  Terminal  Help
ubuntu@domU-12-31-39-02-88-C1:~$ fastqc -h

          FastQC - A high throughput sequence QC analysis tool

SYNOPSIS

      fastqc seqfile1 seqfile2 .. seqfileN

   fastqc [-o output dir] [--{no}extract] [-f fastq|bam|sam]
          [-c contaminant file] seqfile1 .. seqfileN

DESCRIPTION

   FastQC reads a set of sequence files and produces from each one a quality
   control report consisting of a number of different modules, each one of
   which will help to identify a different potential type of problem in your
   data.

   If no files to process are specified on the command line then the program
   will start as an interactive graphical application.  If files are provided
   on the command line then the program will run with no user interaction
   required.  In this mode it is suitable for inclusion into a standardised
   analysis pipeline.

   The options for the program as as follows:

   -h --help       Print this help file and exit

   -v --version    Print the version of the program and exit

   -o --outdir     Create all output files in the specified output directory.
                   Please note that this directory must exist as the program
                   will not create it.  If this option is not set then the
                   output file for each sequence file is created in the same
```

Computer

OR

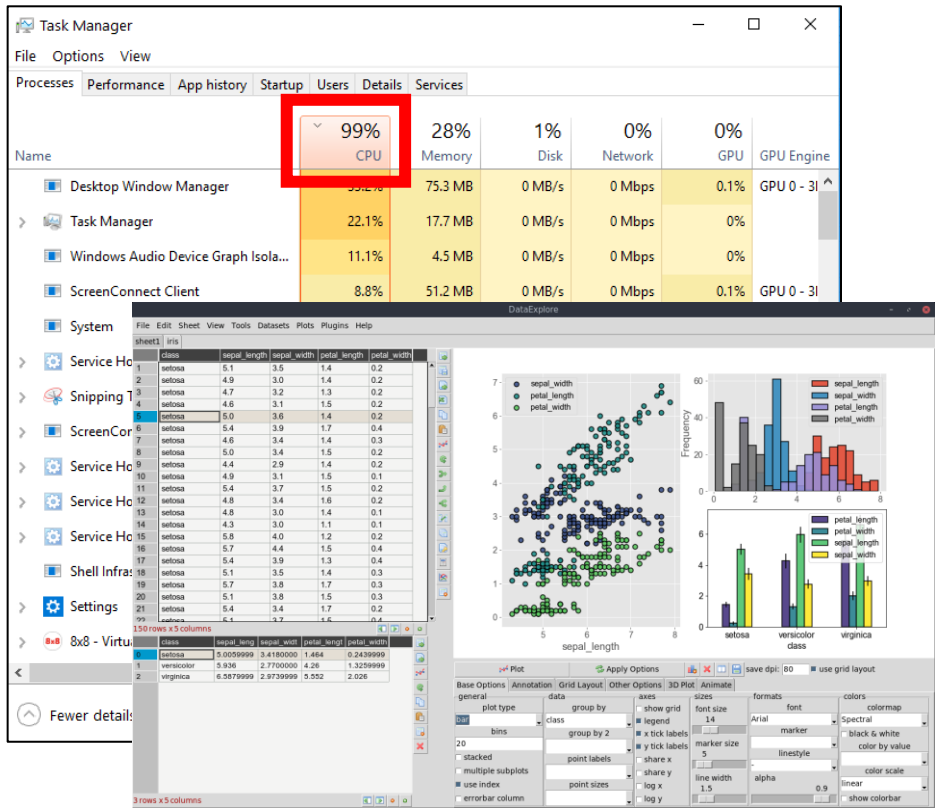
https://en.wikipedia.org/wiki/Supercomputer

Result



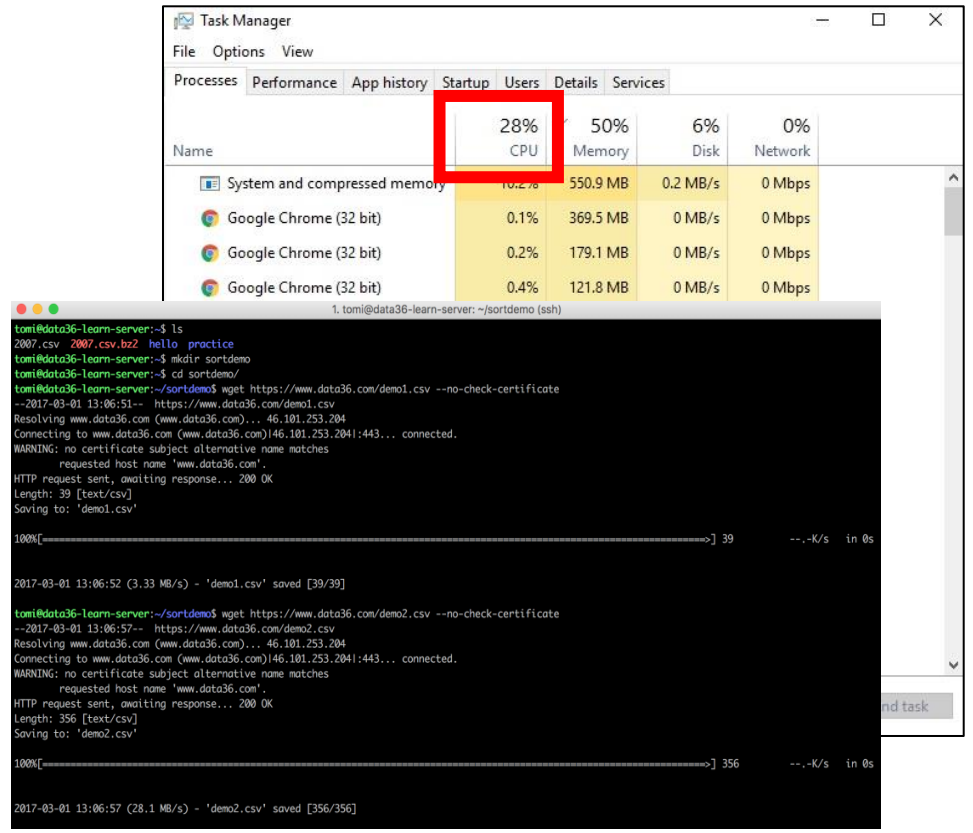
https://du-bii.github.io/module-5-Methodes-Outils/seance1_NGS/slides.html#1

Why use a command line interface (CLI)?



https://decisionstats.com/2015/12/25/interview-damien-farrell-python-gui-dataexplore-python-rstats-pydata/

VS



https://data36.com/command-line-data-science-introduction-to-bash/

Click to go back to the beginning

LEARNING OBJECTIVES

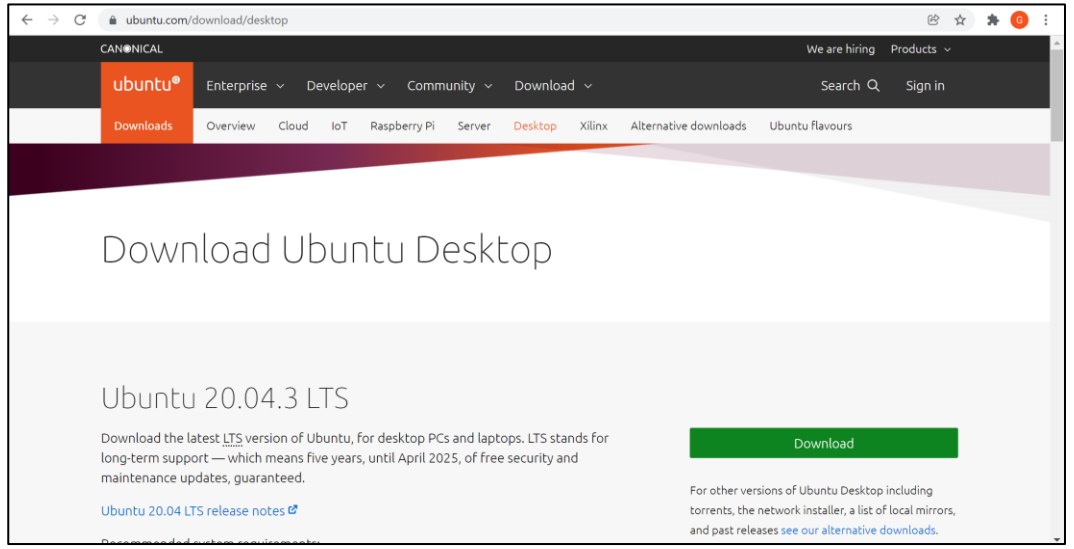
1. *Explain bioinformatics*
2. *Describe the application of bioinformatics in health science research*
3. *Describe the benefits of using a command line interface (CLI) over a graphical user interface (GUI)*
4. ***Install BASH terminal and RStudio***

Virtual Machine (VM)

Follow this guide to install and set-up Ubuntu VM: https://itsfoss.com/install-linux-in-virtualbox/

1. **Download VM by clicking the correct host https://www.virtualbox.org/wiki/Downloads**

2. **Download Ubuntu iso https://ubuntu.com/desktop**



Oracle VirtualBox and Ubuntu

Demonstration

Windows Subsystem for Linux (WSL)

To use BASH on Windows, enable WSL

Advantages:

- Use BASH terminal on Windows as if you are on Linux machine.
- BASH terminal can be accessed using: PowerShell, command prompt or any other terminal programs.
- Access files and run bioinformatics program directly on your computer.

Disadvantages:

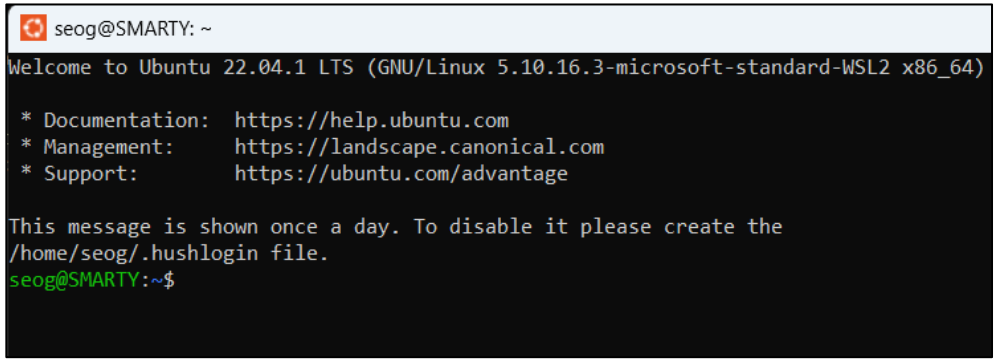
- You can't undo what you just did and there is no trash can.

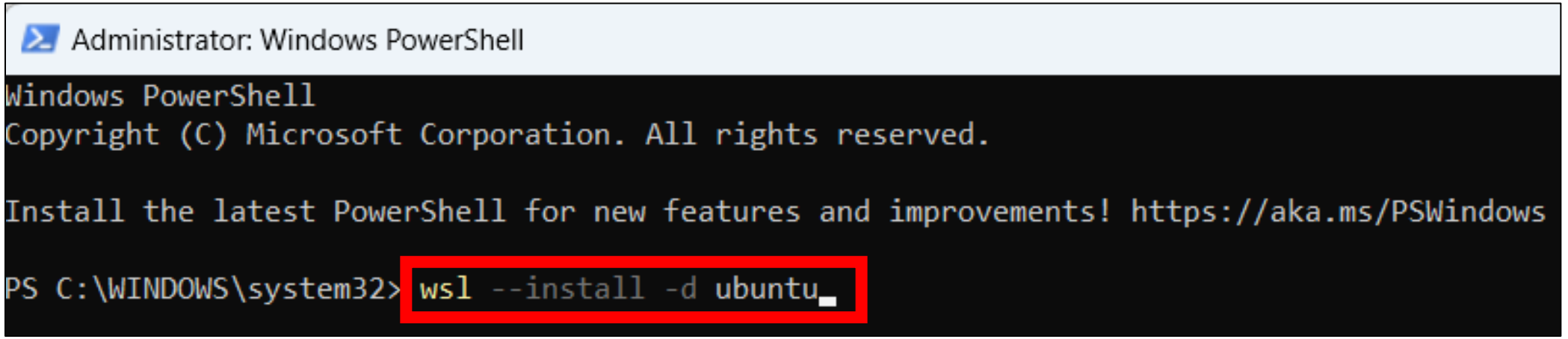
Windows WSL PowerShell

If you have enabled WSL but cannot find Ubuntu on Microsoft Store, use command line to install.

1. **Open PowerShell as administrator**
2. **Type the following command**

> `wsl --install -d ubuntu`





Windows Subsystem for Linux (WSL)

Demonstration

R and Rstudio

Demonstration

Click to go back to the beginning

LEARNING OBJECTIVES

1. *Explain the field of bioinformatics*
2. *Describe the application of bioinformatics in health science research*
3. *Describe the benefits of using a command line interface (CLI) over a graphical user interface (GUI)*
4. *Install BASH terminal and RStudio*

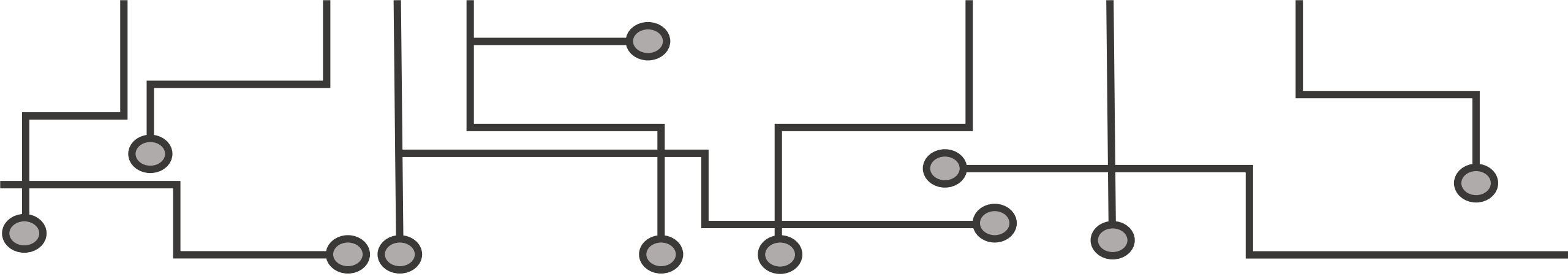
* Items are recommended

HELPFUL RESOURCES

1. ***What is CLI*** *https://www.hostinger.com/tutorials/what-is-cli*
2. ***BASH manual:*** *https://www.gnu.org/software/bash/manual/bash.html*
3. ***Information on Linux folder structure***
   https://www.howtogeek.com/117435/htg-explains-the-linux-directory-structure-explained/
4. ****The BASH Guide:*** *https://guide.bash.academy/*
5. ****Learn Enough Command-Line to be dangerous (free first few chapters):***
   https://www.learnenough.com/command-line-tutorial

YouTube Videos

1. ****Joe Collins - Beginner's Guide to the Bash Terminal:***
   https://www.youtube.com/watch?v=oxuRxtrO2Ag
2. ****Traversy Media – Shell Scripting Crash Course – Beginner Level:***
   https://www.youtube.com/watch?v=v-F3YLd6oMw



THANK YOU FOR ATTENDING!

Please make sure to fill out the Exit Survey at Slido.com with #4156734

We value your feedback!

More questions? Please email us at mmid.bioinformatics.workshop@gmail.com or post them to the workshop slack channel