

The UNIX terminal

```
Dsuite — ssh popgen@ec2-3-93-189-33.compute-1.amazonaws.com — 132x44
...k/Dsuite/Dsuite — -bash  ...Compile/Dsuite — -bash  ...ompile/Dsuite — -bash  ...mpile/Dsuite — -bash  ...ute-1.amazonaws.com +

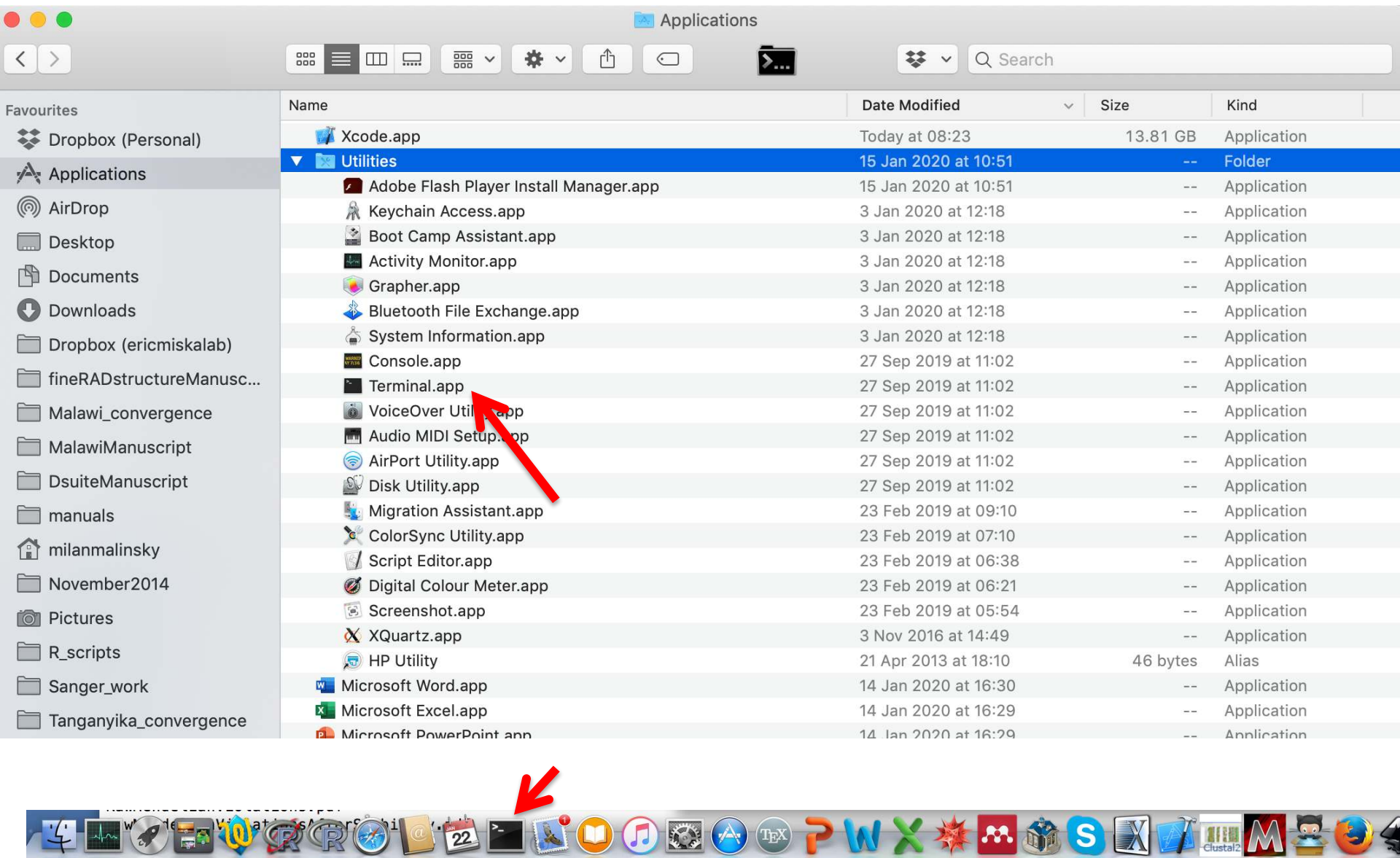
Last login: Sun Jan 19 10:30:59 on ttys015
Milans-MBP-4:Dsuite milanmalinsky$ ssh popgen@ec2-3-93-189-33.compute-1.amazonaws.com
The authenticity of host 'ec2-3-93-189-33.compute-1.amazonaws.com (3.93.189.33)' can't be established.
ECDSA key fingerprint is SHA256:G3fd+zxMfb2dzHBU/luHuqsq+PmeaeAfvUzk/WX2Suw.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'ec2-3-93-189-33.compute-1.amazonaws.com,3.93.189.33' (ECDSA) to the list of known hosts.
popgen@ec2-3-93-189-33.compute-1.amazonaws.com's password:

#####
##          Workshop on Popgen 2020          ##
##          Cesky Krumlov                    ##
##          @evomics #wpsg2020              ##
#####

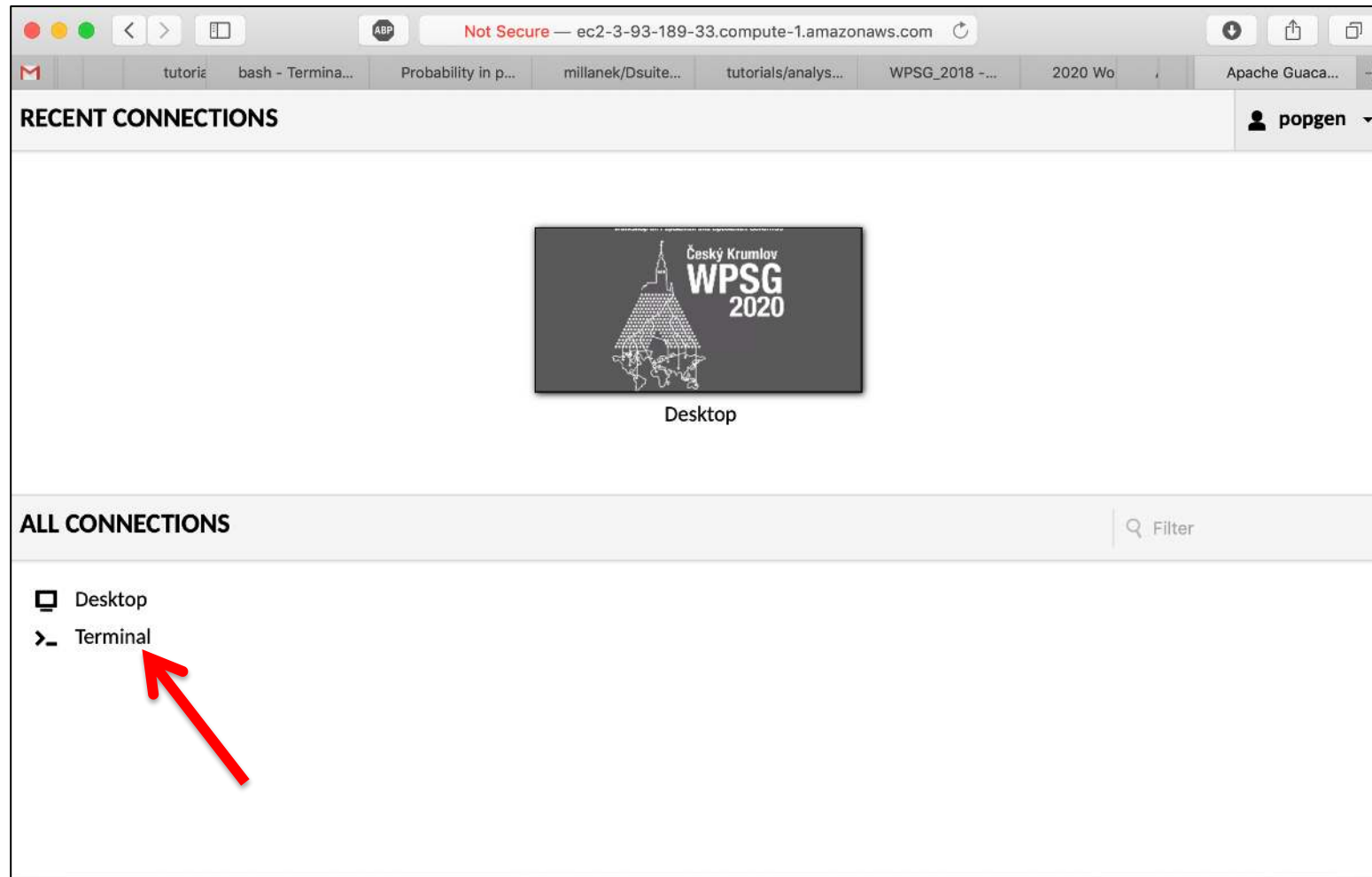
Welcome to Ubuntu 19.10 (5.3.0-1009-aws).

Last login: Sun Jan 19 00:26:57 2020 from 90.178.205.213
(base) 10:54:08 ~$ ls
Desktop  Downloads  Music  Public  Templates  igv  software  workshop_materials
Documents  Miniconda3-latest-Linux-x86_64.sh  Pictures  R  Videos  miniconda3  thinclient_drives
(base) 10:54:11 ~$ cd workshop_materials/
(base) 10:56:39 ~/workshop_materials$ ls -lh
total 60K
drwxr-xr-x 2 popgen workshop 4.0K Jan 18 18:09 20_R_intro
drwxr-xr-x 3 popgen workshop 4.0K Jan 18 18:41 20_python_jupyter_pandas
drwxr-xr-x 3 popgen workshop 4.0K Jan 17 15:57 20_unix_intro
drwxr-xr-x 2 popgen workshop 4.0K Jan 18 21:46 21_first_steps_genomics
drwxr-xr-x 2 popgen workshop 4.0K Jan 17 15:04 21_probability
drwxr-xr-x 4 popgen workshop 4.0K Jan 15 19:21 22_fastsimcoal
drwxr-xr-x 4 popgen workshop 4.0K Jan 17 15:28 23_environmental_selection
drwxr-xr-x 2 popgen workshop 4.0K Jan 18 19:02 24_ARG_Relate
drwxr-xr-x 8 popgen workshop 4.0K Jan 18 19:15 25_mutation_spectrum
drwxr-xr-x 8 popgen workshop 4.0K Jan 17 18:09 27_demography_PSMC_MSMC
drwxr-xr-x 4 popgen workshop 4.0K Jan 17 16:06 27_population_structure
drwxr-xr-x 4 popgen workshop 4.0K Jan 18 12:17 28_species_tree_inference
drwxr-xr-x 2 popgen workshop 4.0K Jan 17 18:36 29_Dsuite_introgession
drwxr-xr-x 2 popgen workshop 4.0K Jan 17 18:30 29_topology_weighting
drwxr-xr-x 8 popgen workshop 4.0K Jan 18 21:50 mushi-1.0-alpha
(base) 10:56:41 ~/workshop_materials$
```

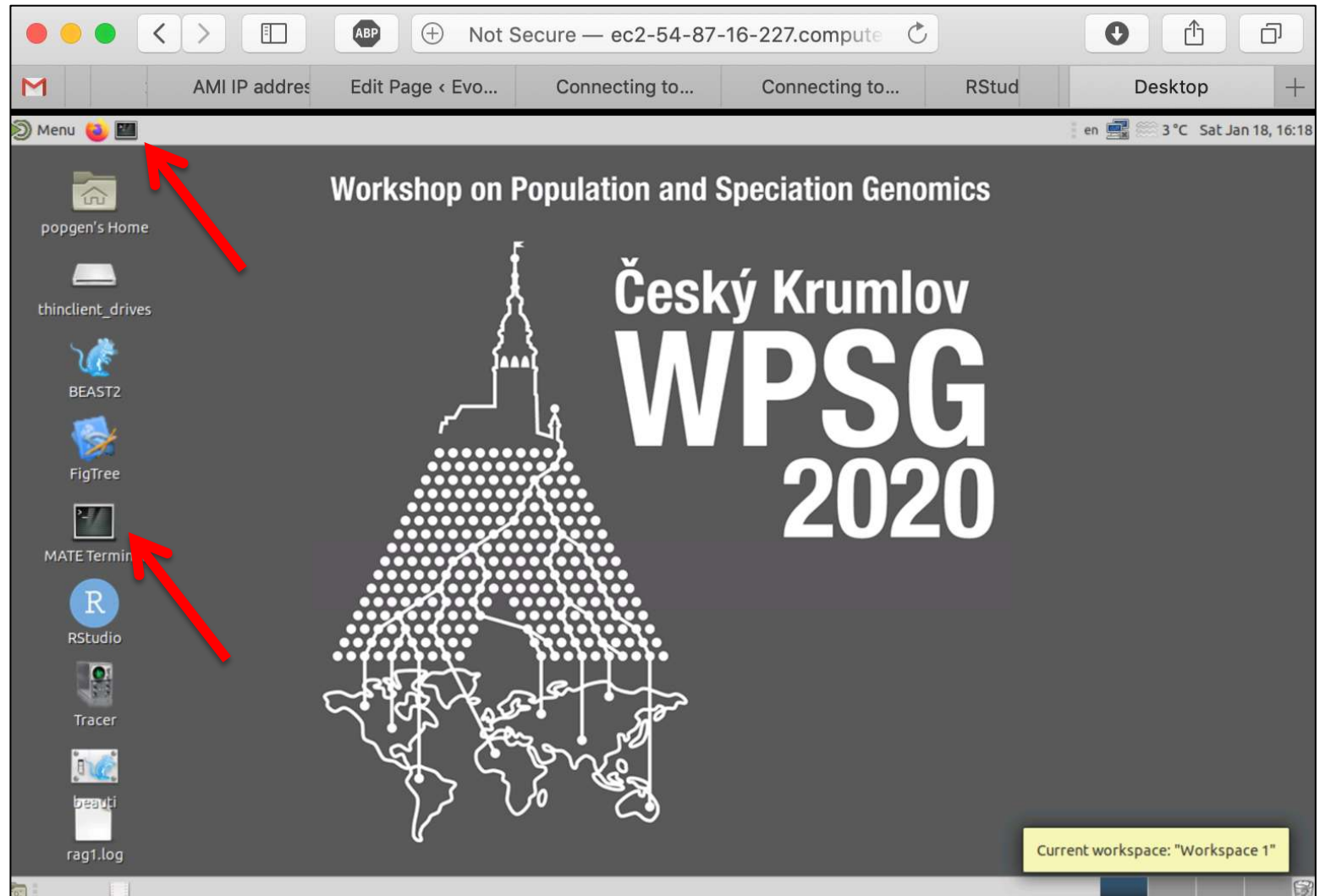
On a Mac computer



On the Guacamole landing page



On the Guacamole desktop



Why use something from the 1960s?

```
Dsuite — ssh popgen@ec2-3-93-189-33.compute-1.amazonaws.com — 132x44
...k/Dsuite/Dsuite — -bash  ...Compile/Dsuite — -bash  ...ompile/Dsuite — -bash  ...mple/Dsuite — -bash  ...ute-1.amazonaws.com +

Last login: Sun Jan 19 10:30:59 on ttys015
Milans-MBP-4:Dsuite milannalinsky$ ssh popgen@ec2-3-93-189-33.compute-1.amazonaws.com
The authenticity of host 'ec2-3-93-189-33.compute-1.amazonaws.com (3.93.189.33)' can't be established.
ECDSA key fingerprint is SHA256:G3fd-zxMF92dzHBU/lUHqsg+PmeaeAfVUZk/WX2Suv.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'ec2-3-93-189-33.compute-1.amazonaws.com,3.93.189.33' (ECDSA) to the list of known hosts.
popgen@ec2-3-93-189-33.compute-1.amazonaws.com's password:

#####
##      Workshop on Popgen 2020      ##
##      Cesky Krumlov               ##
##      @evanils @wpsg2020          ##
#####

Welcome to Ubuntu 19.10 (5.3.0-1009-aws).

Last login: Sun Jan 19 00:26:57 2020 from 90.178.205.213
(base) 10:54:08 ~$ ls
Desktop  Downloads  Music  Public  Templates  igv  software  workshop_materials
Documents  Miniconda3-latest-Linux-x86_64.sh  Pictures  R  Videos  miniconda3  thinclient_drives
(base) 10:54:11 ~$ cd workshop_materials/
(base) 10:56:39 ~/workshop_materials$ ls -lh
total 60K
drwxr-xr-x 2 popgen workshop 4.0K Jan 18 18:09 20_R_intro
drwxr-xr-x 3 popgen workshop 4.0K Jan 18 18:41 20_python_jupyter_pandas
drwxr-xr-x 3 popgen workshop 4.0K Jan 17 15:37 20_unix_intro
drwxr-xr-x 2 popgen workshop 4.0K Jan 18 21:46 21_first_steps_genomics
drwxr-xr-x 2 popgen workshop 4.0K Jan 17 15:04 21_probability
drwxr-xr-x 4 popgen workshop 4.0K Jan 15 19:21 22_fastsimcoal
drwxr-xr-x 4 popgen workshop 4.0K Jan 17 15:28 23_environmental_selection
drwxr-xr-x 2 popgen workshop 4.0K Jan 18 19:02 24_ARG_Relate
drwxr-xr-x 8 popgen workshop 4.0K Jan 18 19:15 25_mutation_spectrum
drwxr-xr-x 8 popgen workshop 4.0K Jan 17 18:09 27_denography_PSMC_MSMC
drwxr-xr-x 4 popgen workshop 4.0K Jan 17 16:06 27_population_structure
drwxr-xr-x 4 popgen workshop 4.0K Jan 18 12:17 28_species_tree_inference
drwxr-xr-x 2 popgen workshop 4.0K Jan 17 18:36 29_Dsuite_introgression
drwxr-xr-x 2 popgen workshop 4.0K Jan 17 18:30 29_topology_weighting
drwxr-xr-x 8 popgen workshop 4.0K Jan 18 21:50 nushi-1.0-alpha
(base) 10:56:41 ~/workshop_materials$
```



Why use something from the 1960s?

```
Dsuite — ssh popgen@ec2-3-93-189-33.compute-1.amazonaws.com — 132x44
...k/Dsuite/Dsuite — -bash    ...Compile/Dsuite — -bash    ...ompile/Dsuite — -bash    ...mple/Dsuite — -bash    ...ute-1.amazonaws.com +

Last login: Sun Jan 19 10:30:59 on ttys015
Milans-MBP-4:Dsuite milannalinsky$ ssh popgen@ec2-3-93-189-33.compute-1.amazonaws.com
The authenticity of host 'ec2-3-93-189-33.compute-1.amazonaws.com (3.93.189.33)' can't be established.
ECDSA key fingerprint is SHA256:G3fd+zMF92dzHBU/1UHqsg+PmeaeAfVUZk/WX2Suv.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'ec2-3-93-189-33.compute-1.amazonaws.com,3.93.189.33' (ECDSA) to the list of known hosts.
popgen@ec2-3-93-189-33.compute-1.amazonaws.com's password:

#####
##      Workshop on Popgen 2020      ##
##      Cesky Krumlov              ##
##      @evonics @wpsg2020         ##
#####

Welcome to Ubuntu 19.10 (5.3.0-1009-aws).

Last login: Sun Jan 19 00:26:57 2020 from 90.178.205.213
(base) 10:54:08 ~: $ ls
Desktop  Downloads  Music  Public  Templates  igv  software  workshop_materials
Documents  Miniconda3-latest-Linux-x86_64.sh  Pictures  R  Videos  miniconda3  thinclient_drives
(base) 10:54:11 ~: $ cd workshop_materials/
(base) 10:56:39 ~/workshop_materials $ ls -lh
total 60K
drwxr-xr-x 2 popgen workshop 4.0K Jan 18 18:09 20_R_intro
drwxr-xr-x 3 popgen workshop 4.0K Jan 18 18:41 20_python_jupyter_pandas
drwxr-xr-x 3 popgen workshop 4.0K Jan 18 17:57 20_unix_intro
drwxr-xr-x 2 popgen workshop 4.0K Jan 18 21:46 21_first_steps_genomics
drwxr-xr-x 2 popgen workshop 4.0K Jan 17 15:04 21_probability
drwxr-xr-x 4 popgen workshop 4.0K Jan 15 19:21 22_fastsimcoal
drwxr-xr-x 4 popgen workshop 4.0K Jan 17 15:28 23_environmental_selection
drwxr-xr-x 2 popgen workshop 4.0K Jan 18 19:02 24_ARG_Relate
drwxr-xr-x 8 popgen workshop 4.0K Jan 18 19:15 25_mutation_spectrum
drwxr-xr-x 8 popgen workshop 4.0K Jan 17 18:09 27_denography_PSMC_MSNC
drwxr-xr-x 4 popgen workshop 4.0K Jan 17 16:06 27_population_structure
drwxr-xr-x 4 popgen workshop 4.0K Jan 18 12:17 28_species_tree_inference
drwxr-xr-x 2 popgen workshop 4.0K Jan 17 18:36 29_Dsuite_introgression
drwxr-xr-x 2 popgen workshop 4.0K Jan 17 18:30 29_topology_weighting
drwxr-xr-x 8 popgen workshop 4.0K Jan 18 21:50 nushi-1.0-alpha
(base) 10:56:41 ~/workshop_materials $
```



Scripting: Write down a sequence of commands to perform a task

In genomics, a task almost always takes minutes, sometimes hours - not fun to sit and wait this long for the next mouse-click.

Why use something from the 1960s?

```
Dsuite — ssh popgen@ec2-3-93-189-33.compute-1.amazonaws.com — 132x44
...k/Dsuite/Dsuite — -bash    ...Compile/Dsuite — -bash    ...mple/Dsuite — -bash    ...ute-1.amazonaws.com +

Last login: Sun Jan 19 10:30:59 on ttys015
Milans-MBP-4:Dsuite milannalinsky ssh popgen@ec2-3-93-189-33.compute-1.amazonaws.com
The authenticity of host 'ec2-3-93-189-33.compute-1.amazonaws.com (3.93.189.33)' can't be established.
ECDSA key fingerprint is SHA256:G3fd+zMF92dzHBU/LUHuqsg+PmeaeAFVUZk/WX2Suv.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'ec2-3-93-189-33.compute-1.amazonaws.com,3.93.189.33' (ECDSA) to the list of known hosts.
popgen@ec2-3-93-189-33.compute-1.amazonaws.com's password:

#####
##      Workshop on Popgen 2020      ##
##      Cesky Krumlov                ##
##      @evonics @wpsg2020          ##
#####

Welcome to Ubuntu 19.10 (5.3.0-1009-aws).

Last login: Sun Jan 19 00:26:57 2020 from 90.178.205.213
(base) 10:54:08 ~: $ ls
Desktop  Downloads  Music  Public  Templates  igv  software  workshop_materials
Documents  Miniconda3-latest-Linux-x86_64.sh  Pictures  R  Videos  miniconda3  thinclient_drives
(base) 10:54:11 ~: $ cd workshop_materials/
(base) 10:56:39 ~/workshop_materials $ ls -lh
total 60K
drwxr-xr-x 2 popgen workshop 4.0K Jan 18 18:09 20_R_intro
drwxr-xr-x 3 popgen workshop 4.0K Jan 18 18:41 20_python_jupyter_pandas
drwxr-xr-x 3 popgen workshop 4.0K Jan 17 15:37 20_unix_intro
drwxr-xr-x 2 popgen workshop 4.0K Jan 18 21:46 21_first_steps_genomics
drwxr-xr-x 2 popgen workshop 4.0K Jan 17 15:04 21_probability
drwxr-xr-x 4 popgen workshop 4.0K Jan 15 19:21 22_fastsimcoal
drwxr-xr-x 4 popgen workshop 4.0K Jan 17 15:28 23_environmental_selection
drwxr-xr-x 2 popgen workshop 4.0K Jan 18 19:02 24_ARG_Relate
drwxr-xr-x 8 popgen workshop 4.0K Jan 18 19:15 25_mutation_spectrum
drwxr-xr-x 8 popgen workshop 4.0K Jan 17 18:09 27_denography_PSMC_MSMC
drwxr-xr-x 4 popgen workshop 4.0K Jan 17 16:06 27_population_structure
drwxr-xr-x 4 popgen workshop 4.0K Jan 18 12:17 28_species_tree_inference
drwxr-xr-x 2 popgen workshop 4.0K Jan 17 18:36 29_Dsuite_introgression
drwxr-xr-x 2 popgen workshop 4.0K Jan 17 18:30 29_topology_weighting
drwxr-xr-x 8 popgen workshop 4.0K Jan 18 21:50 nushi-1.0-alpha
(base) 10:56:41 ~/workshop_materials $
```



Scripting: Write down a sequence of commands to perform a task

In genomics, a task almost always takes minutes, sometimes hours - not fun to sit and wait this long for the next mouse-click.

Easy remote access: Running a real world genomics project on your own computer is impossible; you will usually access high performance compute facilities at your university.

Why use something from the 1960s?

```
Dsuite — ssh popgen@ec2-3-93-189-33.compute-1.amazonaws.com — 132x44
...k/Dsuite/Dsuite — -bash ...Compile/Dsuite — -bash ...ompile/Dsuite — -bash ...mple/Dsuite — -bash ...ute-1.amazonaws.com +

Last login: Sun Jan 19 10:30:59 on ttys015
Milans-MBP-4:Dsuite milannalinsky ssh popgen@ec2-3-93-189-33.compute-1.amazonaws.com
The authenticity of host 'ec2-3-93-189-33.compute-1.amazonaws.com (3.93.189.33)' can't be established.
ECDSA key fingerprint is SHA256:G3fd+zMF92dzHBU/LUHuqsg+PmeaeAFVUZk/WX2Suv.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'ec2-3-93-189-33.compute-1.amazonaws.com,3.93.189.33' (ECDSA) to the list of known hosts.
popgen@ec2-3-93-189-33.compute-1.amazonaws.com's password:

#####
##      Workshop on Popgen 2020      ##
##      Cesky Krumlov               ##
##      @evonics @wpsg2020          ##
#####

Welcome to Ubuntu 19.10 (5.3.0-1009-aws).

Last login: Sun Jan 19 00:26:57 2020 from 90.178.205.213
(base) 10:54:08 ~$ ls
Desktop  Downloads  Music  Public  Templates  igv  software  workshop_materials
Documents  Miniconda3-latest-Linux-x86_64.sh  Pictures  R  Videos  miniconda3  thinclient_drives
(base) 10:54:11 ~$ cd workshop_materials/
(base) 10:56:39 ~/workshop_materials$ ls -lh
total 60K
drwxr-xr-x 2 popgen workshop 4.0K Jan 18 18:09 20_R_intro
drwxr-xr-x 3 popgen workshop 4.0K Jan 18 18:41 20_python_jupyter_pandas
drwxr-xr-x 3 popgen workshop 4.0K Jan 17 15:37 20_unix_intro
drwxr-xr-x 2 popgen workshop 4.0K Jan 18 21:46 21_first_steps_genomics
drwxr-xr-x 2 popgen workshop 4.0K Jan 17 15:04 21_probability
drwxr-xr-x 4 popgen workshop 4.0K Jan 15 19:21 22_fastsimcoal
drwxr-xr-x 4 popgen workshop 4.0K Jan 17 15:28 23_environmental_selection
drwxr-xr-x 2 popgen workshop 4.0K Jan 18 19:02 24_ARG_Relate
drwxr-xr-x 8 popgen workshop 4.0K Jan 18 19:15 25_mutation_spectrum
drwxr-xr-x 8 popgen workshop 4.0K Jan 17 18:09 27_demography_PSMC_MSMC
drwxr-xr-x 4 popgen workshop 4.0K Jan 17 16:06 27_population_structure
drwxr-xr-x 4 popgen workshop 4.0K Jan 18 12:17 28_species_tree_inference
drwxr-xr-x 2 popgen workshop 4.0K Jan 17 18:36 29_Dsuite_introgression
drwxr-xr-x 2 popgen workshop 4.0K Jan 17 18:30 29_topology_weighting
drwxr-xr-x 8 popgen workshop 4.0K Jan 18 21:50 nushi-1.0-alpha
(base) 10:56:41 ~/workshop_materials$
```



Scripting: Write down a sequence of commands to perform a task

In genomics, a task almost always takes minutes, sometimes hours - not fun to sit and wait this long for the next mouse-click.

Easy remote access: Running a real world genomics project on your own computer is impossible; you will usually access high performance compute facilities at your university.

GUI for many programs not available: Genomics is a fast moving field and developing a graphical interface takes time and effort

Why use something from the 1960s?

```

Dsuite — ssh popgen@ec2-3-93-189-33.compute-1.amazonaws.com — 132x44
...k/Dsuite/Dsuite — -bash ...Compile/Dsuite — -bash ...mpile/Dsuite — -bash ...ute-1.amazonaws.com +
Last login: Sun Jan 19 10:30:59 on ttys015
Milans-MP-4:Dsuite milannalinsky@ ssh popgen@ec2-3-93-189-33.compute-1.amazonaws.com
The authenticity of host 'ec2-3-93-189-33.compute-1.amazonaws.com [3.93.189.33]' can't be established.
ECDSA key fingerprint is SHA256:G3JfDzXMFb2dzHBU/LUHqsg+PmeaefVUZk/WX2Suv.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'ec2-3-93-189-33.compute-1.amazonaws.com,3.93.189.33' (ECDSA) to the list of known hosts.
popgen@ec2-3-93-189-33.compute-1.amazonaws.com's password:

#####
##      Workshop on Popgen 2020      ##
##      Cesky Krumlov              ##
##      @evonics @wpsg2020        ##
#####

Welcome to Ubuntu 19.10 (5.3.0-1009-aws).

Last login: Sun Jan 19 00:26:57 2020 from 90.178.205.213
(base) 10:54:08 ~: $ ls
Desktop  Downloads  Music  Public  Templates  igv  software  workshop_materials
Documents  Miniconda3-latest-Linux-x86_64.sh  Pictures  R  Videos  miniconda3  thinclient_drives
(base) 10:54:11 ~: $ cd workshop_materials/
(base) 10:56:39 ~/workshop_materials $ ls -lh
total 60K
drwxr-xr-x 2 popgen workshop 4.0K Jan 18 18:09 20_R_intro
drwxr-xr-x 3 popgen workshop 4.0K Jan 18 18:41 20_python_jupyter_pandas
drwxr-xr-x 3 popgen workshop 4.0K Jan 17 15:37 20_unix_intro
drwxr-xr-x 2 popgen workshop 4.0K Jan 18 21:46 21_first_steps_genomics
drwxr-xr-x 2 popgen workshop 4.0K Jan 17 15:04 21_probability
drwxr-xr-x 4 popgen workshop 4.0K Jan 15 19:21 22_fastsimcoal
drwxr-xr-x 4 popgen workshop 4.0K Jan 17 15:28 23_environmental_selection
drwxr-xr-x 4 popgen workshop 4.0K Jan 18 19:02 24_ARG_Relate
drwxr-xr-x 8 popgen workshop 4.0K Jan 18 19:15 25_mutation_spectrum
drwxr-xr-x 8 popgen workshop 4.0K Jan 17 18:09 27_demography_PSMC_MSMC
drwxr-xr-x 4 popgen workshop 4.0K Jan 17 16:06 27_population_structure
drwxr-xr-x 4 popgen workshop 4.0K Jan 18 12:17 28_species_tree_inference
drwxr-xr-x 2 popgen workshop 4.0K Jan 17 18:36 29_Dsuite_introgression
drwxr-xr-x 2 popgen workshop 4.0K Jan 17 18:36 29_topology_weighting
drwxr-xr-x 8 popgen workshop 4.0K Jan 18 21:50 nushi-1.0-alpha
(base) 10:56:41 ~/workshop_materials $
```



Scripting: Write down a sequence of commands to perform a task

In genomics, a task almost always takes minutes, sometimes hours - not fun to sit and wait this long for the next mouse-click.

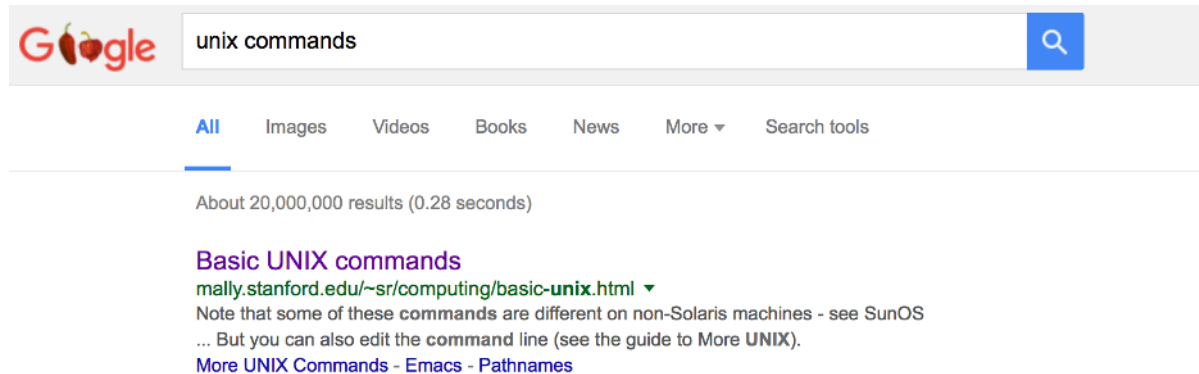
Easy remote access: Running a real world genomics project on your own computer is impossible; you will usually access high performance compute facilities at your university.

GUI for many programs not available: Genomics is a fast moving field and developing a graphical interface takes time and effort

Powerful tools available in UNIX: enabling you to work through large amounts of files, data, and tasks quickly in an automated (programmatic) way

How to survive without 'seeing' your data, without pop-out menus, etc.?

1. **Google:** get a cheat-sheet with a list of possible commands:



2. **Manual pages for each command:** (base) 12:34:35 :~ \$ man ls



What do the `cp`, `vi`, and `awk` commands do? `man cp`, `man vi`, `man awk`

How to survive without 'seeing' your data, without pop-out menus, etc.?

3. **Help:** more concise than manual pages; describes different ways to run the program:

or

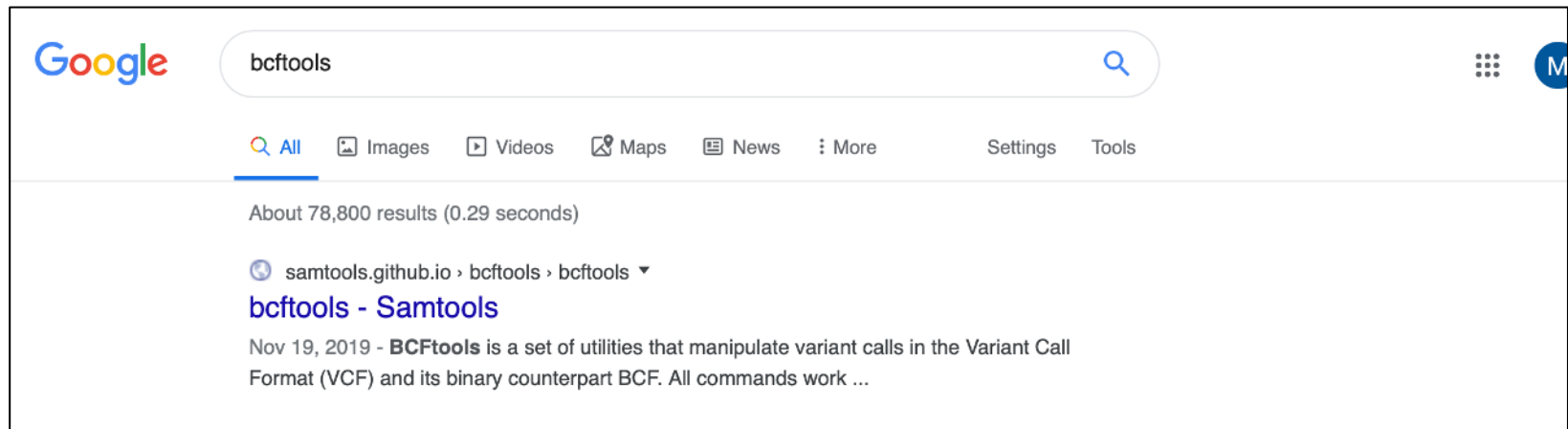
```
(base) 12:37:47 :~ $ bcftools --help
```

```
(base) 12:36:58 :~ $ bcftools -h
```

4. **Genomics tools also sometimes have man pages:**

```
(base) 12:40:32 :~ $ man bcftools
```

5. **Google again:**



A little example

What is a tar archive?



tar = tape archive

When you download programs for UNIX, they often come in tar files.

Genomics data is normally stored compressed to save on disk space/costs.

Example tar commands:

tar -cvzf	creates a compressed tar folder
tar -xvf	extracts an uncompressed tar folder
tar -xvzf	extracts a compressed tar folder

A little example

What is a tar archive?



tar = tape archive

When you download programs for UNIX, they often come in tar files.

Genomics data is normally stored compressed to save on disk space/costs.

