Computers in Science & The Command Line

COMP 364 - Lecture #1
January 4, 2010 - updated January 2012
Derek Ruths

Last class

- Showed how to connect to the comp364 machine
  - Using PuTTY for Windows, Terminal for Mac/Linux
  - host name : `comp364.cs.mcgill.ca`
  - enter `username` and `password`
  - basic commands like ls, cp and cd
  - editing text files with pico

4 “Gears” of Computer Usage

Application usage

Scripting

Algorithms

Application development

4 “Gears” of Computer Usage

Application usage

Scripting

Algorithms

Application development

4 “Gears” of Computer Usage

Application usage

Scripting

Algorithms

Application development

4 “Gears” of Computer Usage

Application usage

Scripting

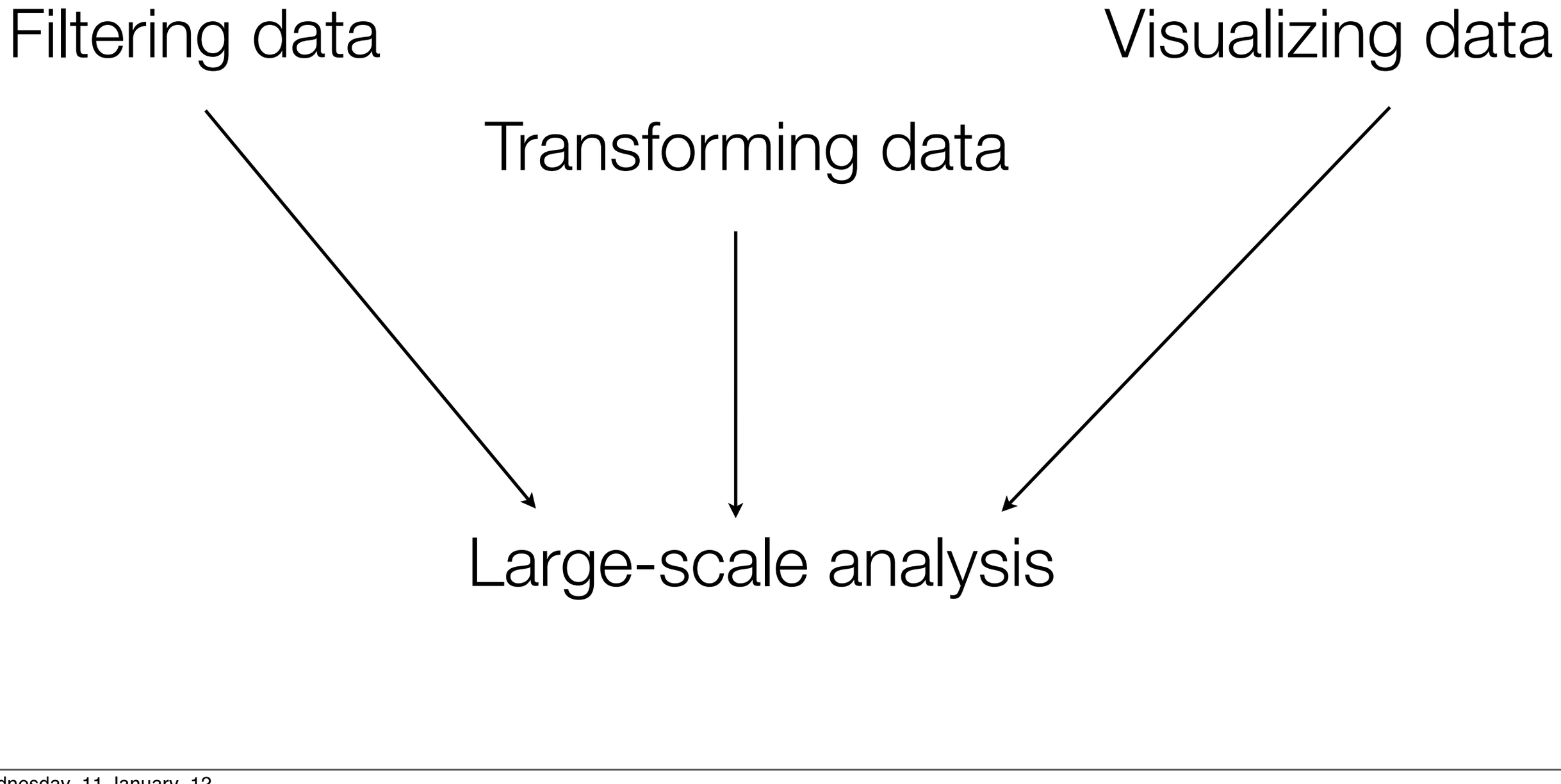
Algorithms

Computer science

Application development

Where does scripting get scientists?

Scripts: VERY short, custom applications



Filtering data

Idea: extract “interesting” data from a dataset

All genes in the human genome with >1000 bp

All data points within 5 standard deviations of the mean

Transforming data

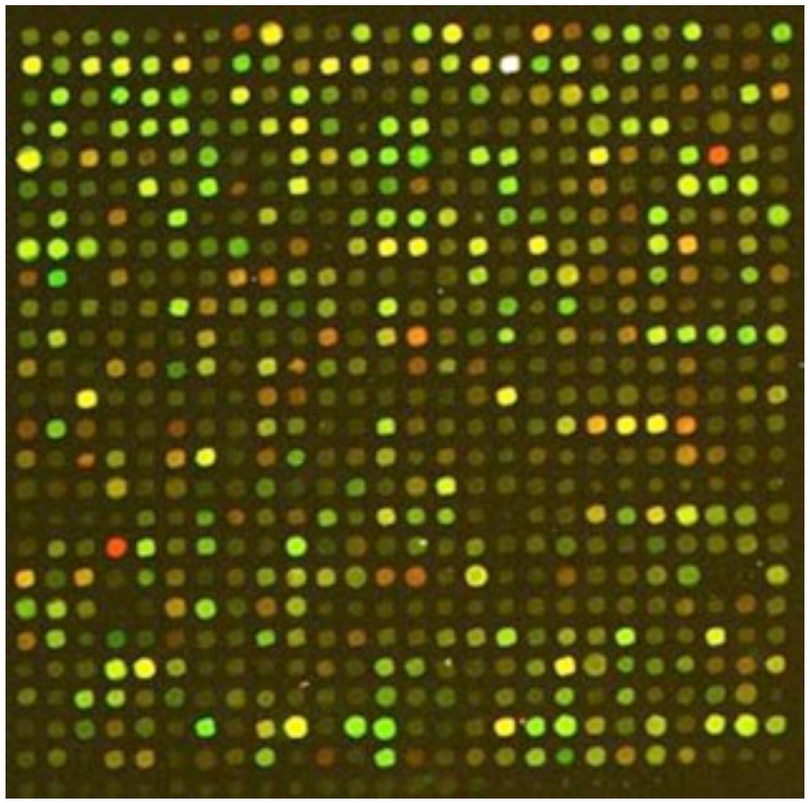
Idea: transforming data from one form to another

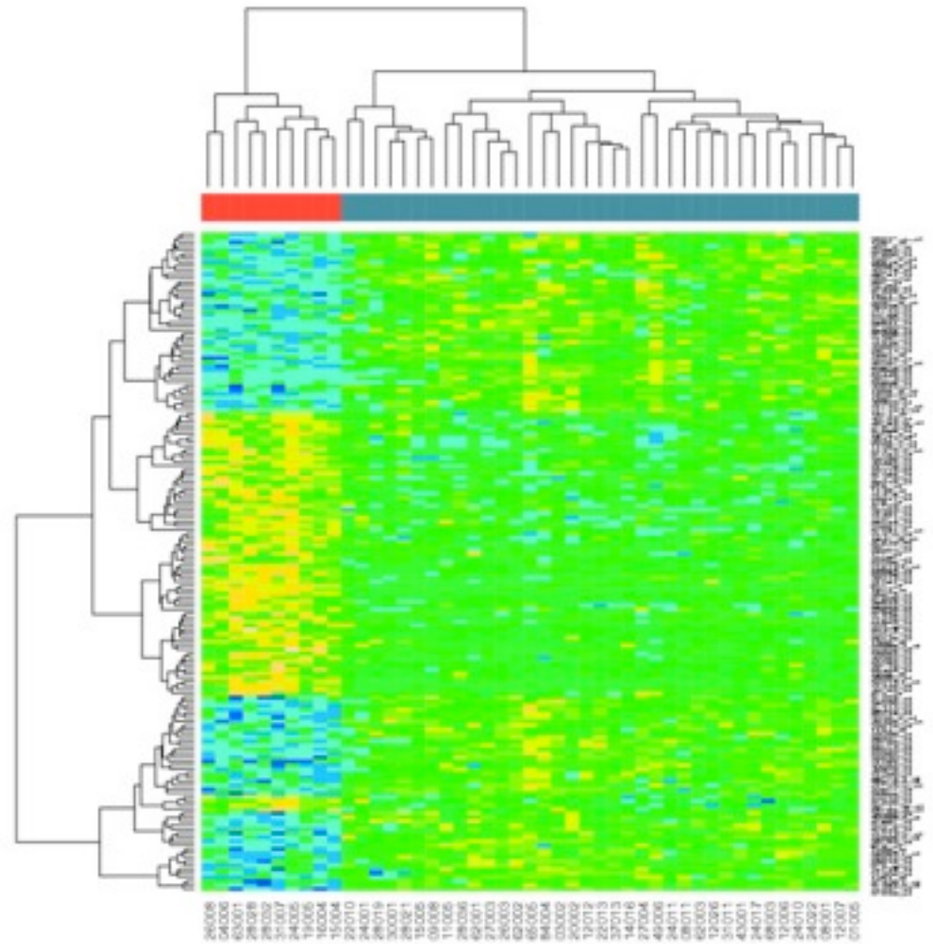
List of genes → List of gene *lengths*

XML → Excel

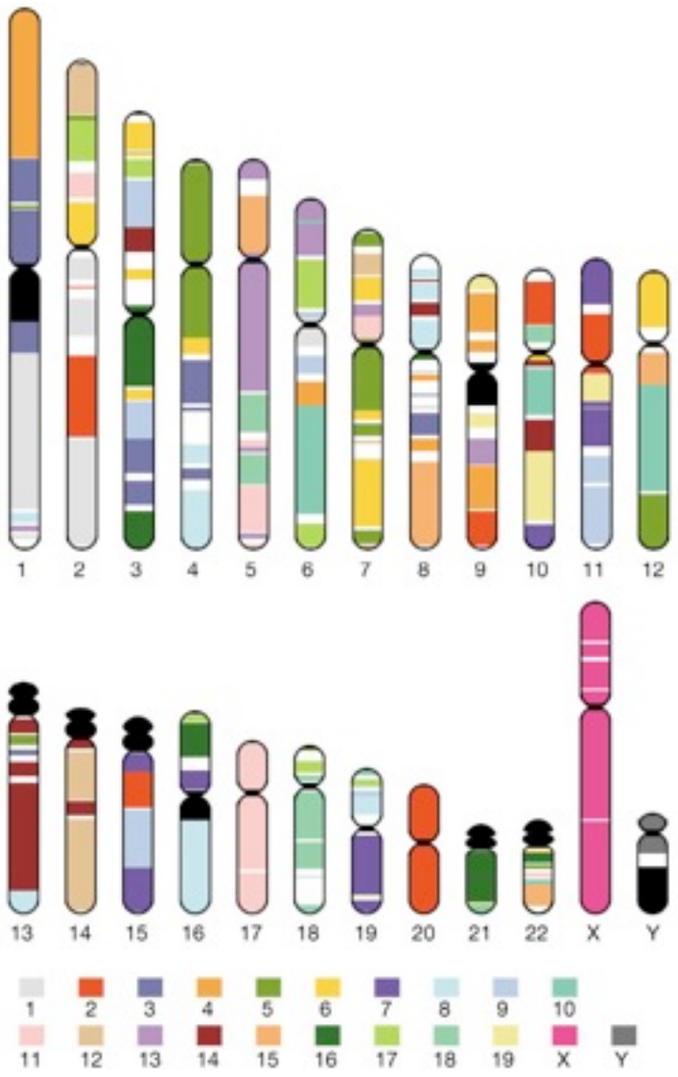
Visualizing data

Idea: quickly make useful plots of data





Idea: putting filtering, transformation, and visualization together



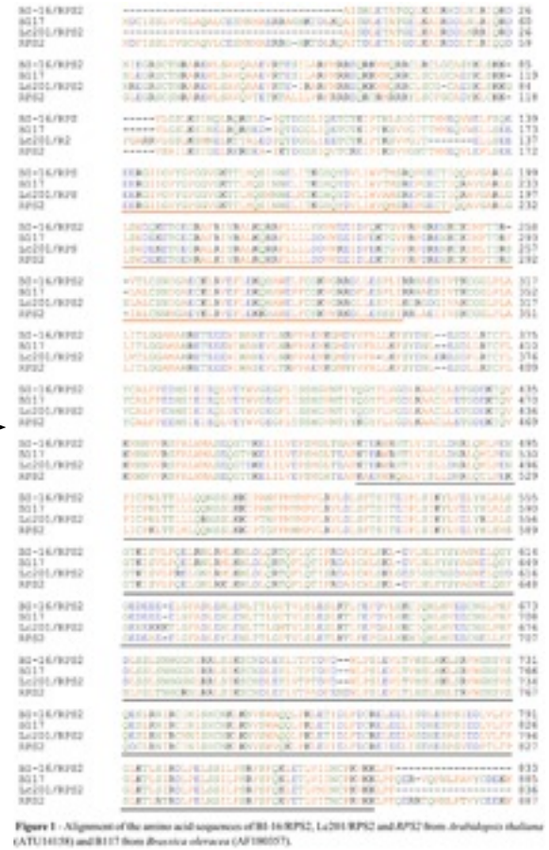
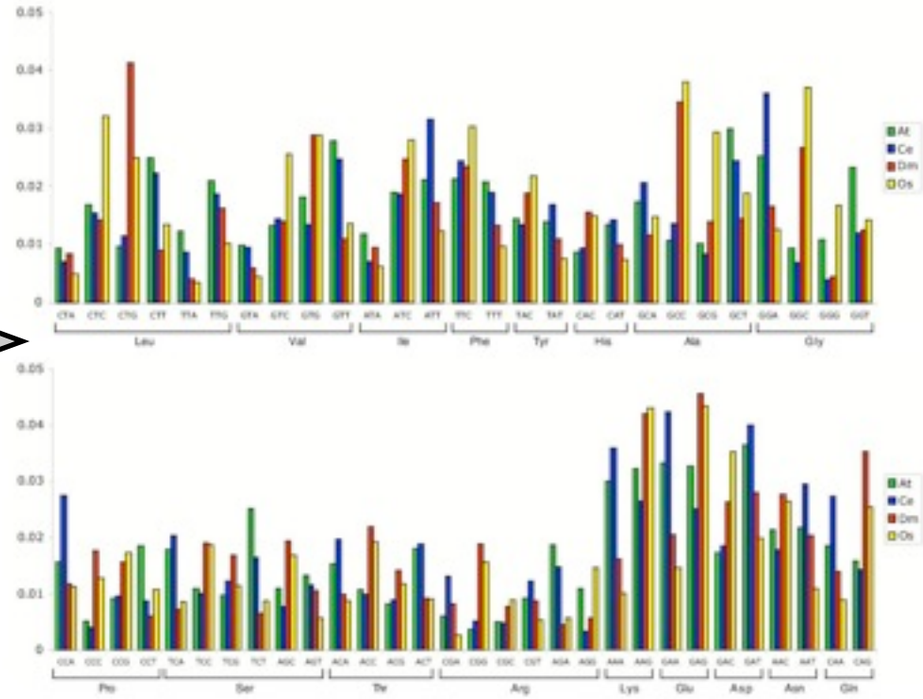


Figure 1 - Alignment of the amino acid sequences of Bl-16/RPS2, Lc201/RPS2 and AtRPS2 from *Arabidopsis thaliana* (ATU14158) and Bl17 from *Brassica oleracea* (AF190057).



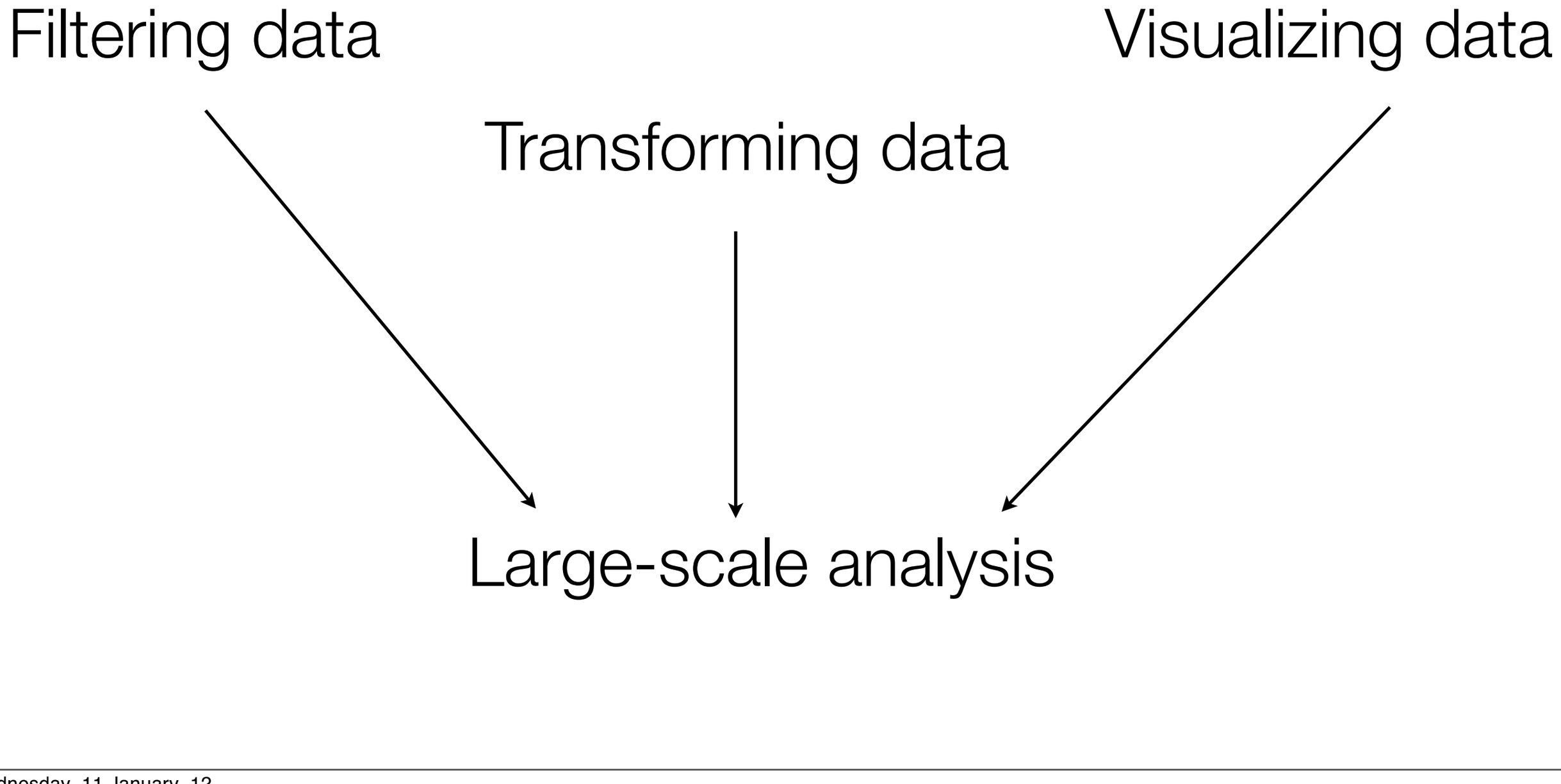
Human genome

Genes of interest

Codon frequencies

Where does scripting get scientists?

These skills can save hundreds of hours.



The Command Line

Communicating with the computer

or

The art of writing one-line programs

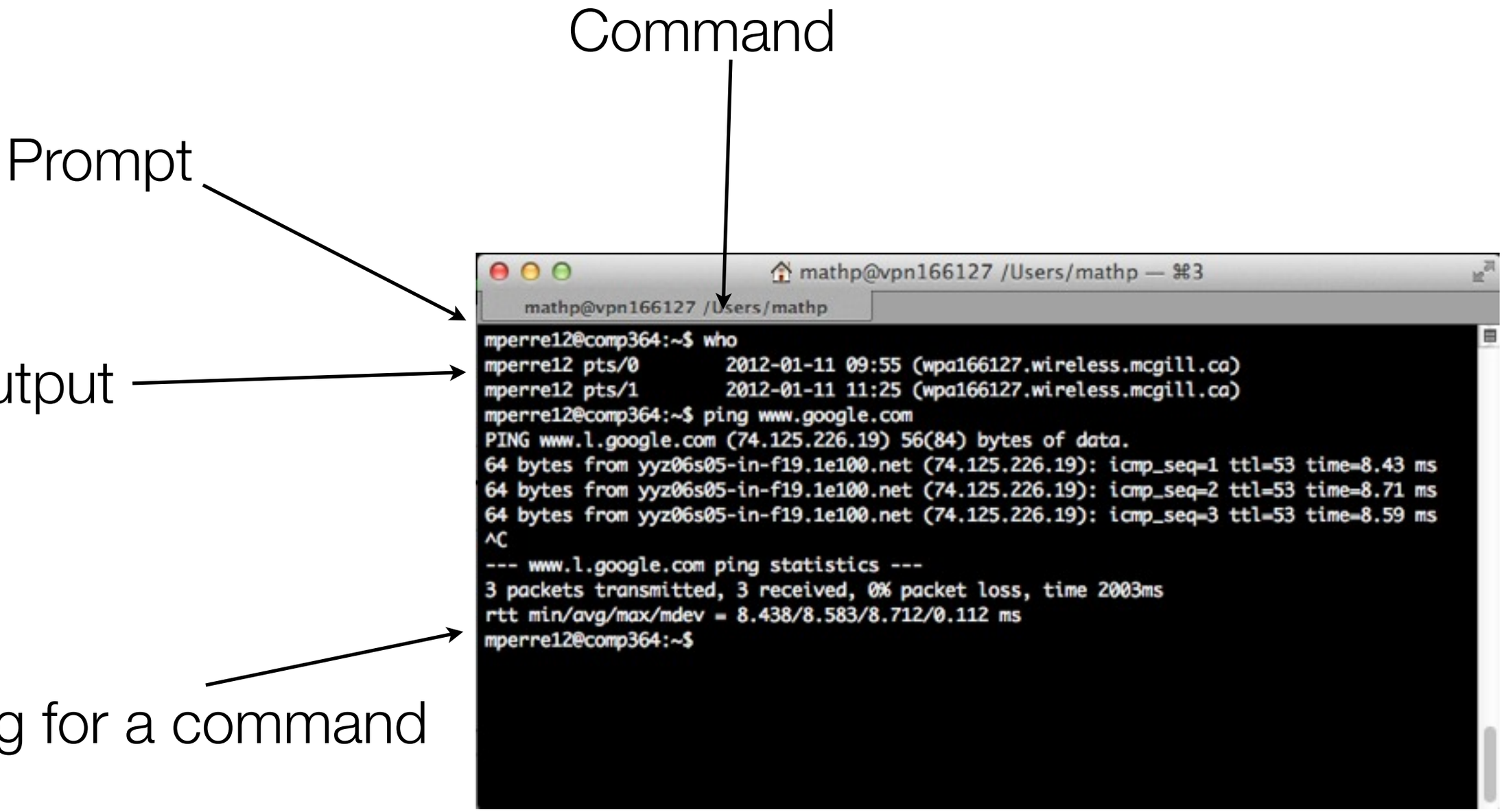
(Constructive) Communication with computers

- **Get information**
  - What is the size of a file?
  - How many words are in a document?
  - How many times is “E. coli” mentioned in a file?
- **Perform a task**
  - Rename a file
  - Download a data set
  - Generate an image
  - Find all the palindromic words in the English language

Every communication is a task (command)

- *What is the size of a file?*
  - Count the number of bytes in a file
- *How many words are in a document?*
  - Count the number of words in a document
- *How many times is “E. coli” mentioned in a file?*
  - Count the number of times “E. coli” is mentioned in a file

Speaking through the command line



Translating commands to the command line

- *What is the size of foo.txt?*
  - Count the number of bytes in foo.txt
  - *ls -l foo.txt*
- *How many words are in foo.txt?*
  - Count the number of words in foo.txt
  - *wc -w foo.txt*
- *How many times is “E. coli” mentioned in foo.txt?*
  - Count the number of times “E. coli” is mentioned in foo.txt
  - *grep -c “E. coli” foo.txt*
- *Rename file foo.txt to bar.txt*
  - *mv foo.txt bar.txt*

The structure of commands

<command> <options/flags> <arguments>

ls -l foo.txt

wc -w foo.txt

mv foo.txt bar.txt

Options

Options always preceded by a “-”

```
ls -l foo.txt

wc -w foo.txt
```

An option can take one or more arguments

```
head -n 3 foo.txt
```

Commands, options, & arguments with spaces

Terms with spaces must be in double quotes

grep -c E. coli foo.txt 

grep -c “E. coli” foo.txt 

One line programs

A single line on the command line can do a lot

In the coming weeks:

- Managing & manipulating files & directories
- Searching and filtering files
- Accessing and downloading internet content