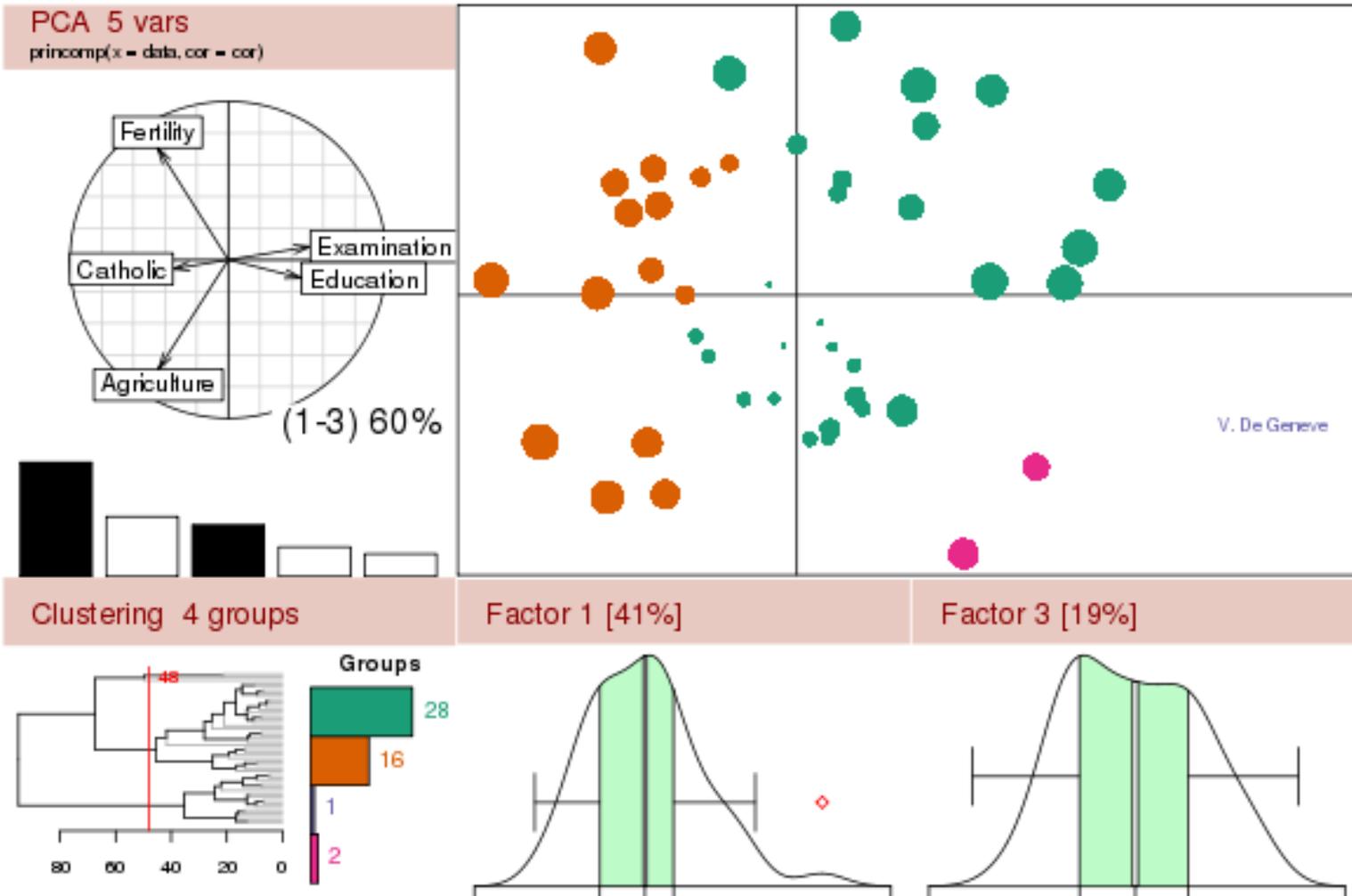


# The R Project



# Avant de commencer

```
source("http://bioconductor.org/biocLite.R")
```

```
biocLite()
```

```
biocLite("Biostrings")
```

```
biocLite("BSgenome")
```

```
biocLite("BSgenome.Dmelanogaster.UCSC.dm3")
```

```
biocLite("BSgenome.Celegans.UCSC.ce2")
```

# Biostrings

Représentation et manipulation de grandes séquences ou ensemble de séquences dans R/Bioconductor

Extraction d'information

Transformation

Alignement

# Bsgenome

Conteneur de séquences génomiques complètes pour différentes espèces

Séquences stockées dans des objets de type Biostrings

# Conteneurs

**DNA/RNA/AAString**: contient une séquence

**DNA/RNA/AAStringSet**: contient un ensemble de séquences

[ ] est employé pour faire des sous-ensembles

[[ ]] extrait une séquence sous forme de DNA/RNA/AAString

**XStringViews**: contient un ensemble de vue sur la meme séquence

Chaque vue est définie par sa position de début et de fin

length, width, [ ] et [[ ]] sont supportés

**MaskedDNA/RNA/AAString**: contient une séquence masquée, c'est à dire un object DNA/RNA/AAString plus un ensemble de masques

Les masques permettent d'ignorer certaines régions de la séquence

La fonction unmasked renvoie la séquence originale dans un object DNA/RNA/AAString

La fonction masks retourne les masques

# Fonctions

**alphabetFrequency**: renvoie la fréquence pour chacune des bases de la séquence

**reverse/complement/reverseComplement**: retourne et/ou complémente la séquence

**subseq**: renvoie une sous-séquence

**matchPattern**: retourne toutes les occurrences d'un motif à l'intérieur d'une séquence

**vmatchPattern**: retourne toutes les occurrences d'un motif à l'intérieur d'un ensemble de séquences

**matchPDict**: retourne toutes les occurrences de plusieurs motifs à l'intérieur d'une séquence

**vmatchPDict**: retourne toutes les occurrences de plusieurs motifs à l'intérieur d'un ensemble de séquences

**read/write.DNAStringSet**: lit/écrit des fichiers au format FASTA