Class09

> Q1: What percentage of structures in the PDB are solved by X-Ray and Electron Microscopy.

```
pdb<- read.csv("Data Export Summary.csv", row.names=1)
```

let's create a function

```
comma_sum <- function(x) {
  sum(as.numeric(gsub(",","", x)))
}

comma_sum(pdb$X.ray)
```

```
[1] 174797
```

```
comma_sum(pdb$EM)
```

```
[1] 15236
```

```
comma_sum(pdb$Total)
```

```
[1] 204352
```

```
#for xray
comma_sum(pdb$X.ray)/comma_sum(pdb$Total)
```

```
[1] 0.8553721
```

```
#for em
comma_sum(pdb$EM)/comma_sum(pdb$Total)
```

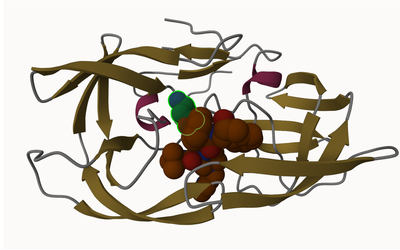
```
[1] 0.07455763
```

> Q2: What proportion of structures in the PDB are protein?

```
round(comma_sum(pdb$Total[1])/comma_sum(pdb$Total), 2)
```

```
[1] 0.87
```

> Q3: Type HIV in the PDB website search box on the home page and determine how many HIV-1 protease structures are in the current PDB?



HIV-PR structure from MERK a bound drug

skipped #Visualizing the HIV-1 protease

> Q4 Water molecules normally have 3 atoms. Why do we see just one atom per water molecule in this structure?

The resolution isnt strong enough to detect the minsulce hydrogen molecules that are attached to the big oxygen molecules. >Q5: There is a critical "conserved" water molecule in the binding site. Can you identify this water molecule? What residue number does this water molecule have

Water molecules bind to ASP25 on each chain # working with structures in R

```
library(bio3d)
hiv <- read.pdb("1hsg")
```

```
  Note: Accessing on-line PDB file
```

```
hiv
```

```

 Call:  read.pdb(file = "1hsg")

   Total Models#: 1
     Total Atoms#: 1686,  XYZs#: 5058  Chains#: 2  (values: A B)

     Protein Atoms#: 1514  (residues/Calpha atoms#: 198)
     Nucleic acid Atoms#: 0  (residues/phosphate atoms#: 0)

     Non-protein/nucleic Atoms#: 172  (residues: 128)
     Non-protein/nucleic resid values: [ HOH (127), MK1 (1) ]

   Protein sequence:
      PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIGGFIKVRQYD
      QILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFPQITLWQRPLVTIKIGGQLKE
      ALLDTGADDTVLEEMSLPGRWKPKMIGGIGGFIKVRQYDQILIEICGHKAIGTVLVGPTP
      VNIIGRNLLTQIGCTLNF

+ attr: atom, xyz, seqres, helix, sheet,
        calpha, remark, call
```

```
attributes(pdb)
```

```
$names
[1] "X.ray"            "EM"               "NMR"              "Multiple.methods"
[5] "Neutron"          "Other"            "Total"

$class
[1] "data.frame"

$row.names
[1] "Protein (only)"          "Protein/Oligosaccharide"
[3] "Protein/NA"              "Nucleic acid (only)"
[5] "Other"                   "Oligosaccharide (only)"
```

```
head(pdb$atom)
```

```
NULL
```

Perform a prediction of flexibility with a technique called NMA (normal mode analysis)

```
adk <- read.pdb("6s36")
```

```
  Note: Accessing on-line PDB file
    PDB has ALT records, taking A only, rm.alt=TRUE
```

```
adk
```

```

 Call:  read.pdb(file = "6s36")

   Total Models#: 1
     Total Atoms#: 1898,  XYZs#: 5694  Chains#: 1  (values: A)

     Protein Atoms#: 1654  (residues/Calpha atoms#: 214)
     Nucleic acid Atoms#: 0  (residues/phosphate atoms#: 0)

     Non-protein/nucleic Atoms#: 244  (residues: 244)
     Non-protein/nucleic resid values: [ CL (3), HOH (238), MG (2), NA (1) ]

   Protein sequence:
      MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVT
      DELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDKI
      VGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIG
      YYSKEAEAGNTKYAKVDGTKPVAEVRADLEKILG

+ attr: atom, xyz, seqres, helix, sheet,
        calpha, remark, call
```

> Q7: How many amino acid residues are there in this pdb object? 198

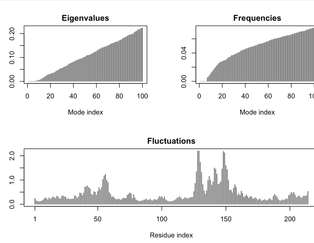
> Q8: Name one of the two non-protein residues? hoh

> Q9: How many protein chains are in this structure? 2

```
# Perform flexiblity prediction
m <- nma(adk)
```

```
 Building Hessian...     Done in 0.029 seconds.
 Diagonalizing Hessian...  Done in 0.306 seconds.
```

```
plot(m)
```



Write out a "movie" of the motion for viewing in Molstar

```
mktrj(m, file="adk_m7.pdb")
```