

Class09

Jiawei Xu

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

	X.ray	EM	NMR	Multiple.methods	Neutron	Other
Protein (only)	152,809	9,421	12,117	191	72	32
Protein/Oligosaccharide	9,008	1,654	32	7	1	0
Protein/NA	8,061	2,944	281	6	0	0
Nucleic acid (only)	2,602	77	1,433	12	2	1
Other	163	9	31	0	0	0
Oligosaccharide (only)	11	0	6	1	0	4
Total						
Protein (only)	174,642					
Protein/Oligosaccharide	10,702					
Protein/NA	11,292					
Nucleic acid (only)	4,127					
Other	203					
Oligosaccharide (only)	22					

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

```
n.xray <- sum(as.numeric(gsub(",", "", pdbstats$X.ray)))
n.em <- sum(as.numeric(gsub(",", "", pdbstats$EM)))
n.total <- sum(as.numeric(gsub(",", "", pdbstats$Total)))

p.xray <- 100* n.xray/n.total
p.em <- (n.em/n.total)*100

round(p.xray, 2)
```

[1] 85.9

```
round(p.em, 2)
```

[1] 7.02

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

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?

