

# Class 9 Project

Garrett Cole

## Introduction to the RCSB Protein Data Bank (PDB)

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

	X-ray	NMR	EM	Multiple.methods	Neutron	Other
Protein (only)	150,417	12,056	8,586	188	72	32
Protein/Oligosaccharide	8,869	32	1,552	6	0	0
Protein/NA	7,943	280	2,690	6	0	0
Nucleic acid (only)	2,522	1,425	74	13	2	1
Other	154	31	6	0	0	0
Oligosaccharide (only)	11	6	0	1	0	4
Total						
Protein (only)	171,351					
Protein/Oligosaccharide	10,459					
Protein/NA	10,919					
Nucleic acid (only)	4,037					
Other	191					
Oligosaccharide (only)	22					

**Question 1: What percentage of structures in the PDB are solved by X-ray and Electron Microscopy?**

$$((169,916+12,908)/196,979) * 100 = 92.81\%$$

**Question 2: What proportion of structures in the PDB are protein?**

$$(171,351/196,979) * 100 = 86.99\%$$

**Question 3: 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?**

646