Class_11

Zaida Rodriguez (PID:A59010549)

11/5/2021

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

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

The Xray is 87.53% and Em is 4.95%

```
library(skimr)
skim(pdb.data)
```

Table 1: Data summary

Name	pdb.data
Number of rows	6
Number of columns	7
Column type frequency:	
numeric	7
Group variables	None

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
X.ray	0	1	26811.83	56749.93	11	703.75	4933.0	8194.00	142419	
NMR	0	1	2254.50	4709.24	6	31.00	152.5	1102.00	11807	
EM	0	1	1515.33	2352.26	0	17.25	525.5	1747.75	6038	
Multiple.methods	0	1	32.33	70.93	0	1.50	4.0	7.25	177	
Neutron	0	1	12.00	28.43	0	0.00	0.0	1.50	70	
Other	0	1	6.17	12.75	0	0.00	0.5	3.25	32	
Total	0	1	30632.17	63786.50	22	1091.50	6635.0	9694.50	160543	

```
round(sum(pdb.data$X.ray) / sum(pdb.data$Total) * 100, 2)
```

```
## [1] 87.53
```

```
round(sum(pdb.data$EM) / sum(pdb.data$Total) * 100, 2)
```

```
## [1] 4.95
```

```
#  to see all the column row percentage:
round((colSums(pdb.data) / sum(pdb.data$Total)) *100, 2)
```

```
##           X.ray              NMR               EM Multiple.methods
##           87.53             7.36             4.95             0.11
##          Neutron            Other            Total
##            0.04             0.02           100.00
```

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

The proportion is 0.87

```
#call for protein only
prot <- pdb.data$Total[1]

#call for the proportion of protein
prot.prop <- prot/sum(pdb.data$Total)
prot.prop
```

```
## [1] 0.873499
```

> 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?

There were 22,824 protease protein structures

2. Visualizing the HIV-1 Protease Structures

See VMD Image

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

We only see one because it can be overwhelming to see the all the atoms separate. Also we know that all 3
atoms are binded together so no need to see them separately

> Q5: There is a conserved water molecule in the binding site. Can you identify this water molecule?
> What residue number does this water molecule have (see note below)?

HOH 308