

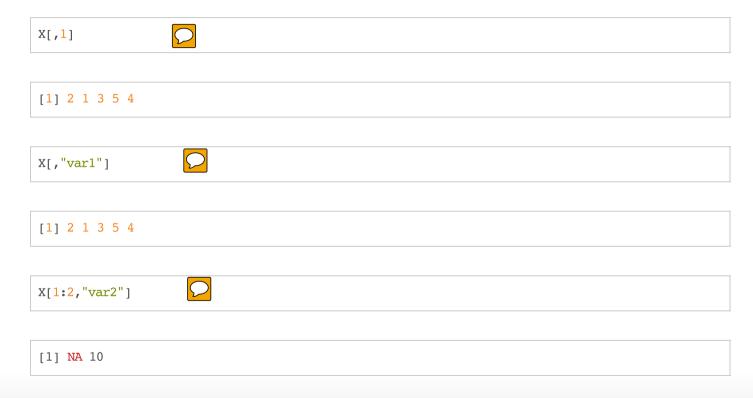
Subsetting and sorting

Jeffrey Leek
Johns Hopkins Bloomberg School of Public Health

Subsetting - quick review

```
set.seed(13435)
X <- data.frame("var1"=sample(1:5), "var2"=sample(6:10), "var3"=sample(11:15))
X <- X[sample(1:5),]; X$var2[c(1,3)] = NA</pre>
X
```

Subsetting - quick review



Logicals ands and ors

```
X[(X$var1 <= 3 & X$var3 > 11),]
```



```
X[(X$var1 <= 3 | X$var3 > 15),]
```

Dealing with missing values

```
X[which(X$var2 > 8),]
```

Sorting

sort(X\$var1) [1] 1 2 3 4 5 sort(X\$var1,decreasing=TRUE) [1] 5 4 3 2 1 sort(X\$var2,na.last=TRUE) [1] 6 9 10 NA NA

Ordering

```
X[order(X$var1),]
```

Ordering

```
X[order(X$var1,X$var3),]
```

Ordering with plyr

```
library(plyr)
arrange(X,var1)
```

```
arrange(X,desc(var1))
```

Adding rows and columns

```
X$var4 <- rnorm(5) X
```

```
var1 var2 var3 var4

1 2 NA 15 0.18760

4 1 10 11 1.78698

2 3 NA 12 0.49669

3 5 6 14 0.06318

5 4 9 13 -0.53613
```

Adding rows and columns

```
Y <- cbind(X,rnorm(5))

Y
```

Notes and further resources

- · R programming in the Data Science Track
- · Andrew Jaffe's lecture notes http://www.biostat.jhsph.edu/~ajaffe/lec_winterR/Lecture%202.pdf