

R Handout

Getting Data Into R

Derek H. Ogle

Always First Command

```
> library(NCStats)
```

Basic Data Frames

```
> setwd("C:/aaaWork/Web/GitHub/NCMTH107/modules/Getting_Data_Into_R")
> dfobj <- read.csv("Iris.csv")
> str(dfobj)
```

```
'data.frame': 150 obs. of 5 variables:
 $ seplen : int  50 46 46 51 55 48 52 49 44 50 ...
 $ sepwid : int  33 34 36 33 35 31 34 36 32 35 ...
 $ petlen : int  14 14 10 17 13 16 14 14 13 16 ...
 $ petwid : int   2 3 2 5 2 2 2 1 2 6 ...
 $ species: Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 1 1 1 1 1 1 ...
```

```
> headtail(dfobj)      # NOT the entire data.frame
```

| | seplen | sepwid | petlen | petwid | species |
|-----|--------|--------|--------|--------|-----------|
| 1 | 50 | 33 | 14 | 2 | setosa |
| 2 | 46 | 34 | 14 | 3 | setosa |
| 3 | 46 | 36 | 10 | 2 | setosa |
| 148 | 57 | 25 | 50 | 20 | virginica |
| 149 | 58 | 28 | 51 | 24 | virginica |
| 150 | 63 | 33 | 60 | 25 | virginica |

Working With Data Frames

```
> dfobj$seplen
```

```
[1] 50 46 46 51 55 48 52 49 44 50 44 47 48 51 48 50 50 43 58 51 49 51 50 46 57 50 54 52
[29] 55 49 54 50 44 47 46 51 50 49 54 54 51 48 48 45 57 51 54 51 52 53 65 62 59 61 60 56
[57] 57 63 70 64 61 55 54 58 55 50 67 56 58 60 57 57 49 56 57 66 52 60 50 55 58 62 59 60
[85] 67 63 56 63 61 64 51 57 61 56 69 55 55 66 68 67 64 67 63 69 65 65 58 68 62 77 67 76
[113] 49 67 59 63 64 79 67 77 63 72 61 61 64 62 77 63 58 72 71 64 60 63 77 60 69 74 56 73
[141] 67 65 69 72 65 64 68 57 58 63
```

```
> dfobj$species
```

| | | | | | | | |
|------|--------|--------|--------|--------|--------|--------|--------|
| [1] | setosa | setosa | setosa | setosa | setosa | setosa | setosa |
| [8] | setosa | setosa | setosa | setosa | setosa | setosa | setosa |
| [15] | setosa | setosa | setosa | setosa | setosa | setosa | setosa |
| [22] | setosa | setosa | setosa | setosa | setosa | setosa | setosa |

```

[29] setosa      setosa      setosa      setosa      setosa      setosa      setosa
[36] setosa      setosa      setosa      setosa      setosa      setosa      setosa
[43] setosa      setosa      setosa      setosa      setosa      setosa      setosa
[50] setosa      versicolor versicolor versicolor versicolor versicolor versicolor
[57] versicolor versicolor versicolor versicolor versicolor versicolor versicolor
[64] versicolor versicolor versicolor versicolor versicolor versicolor versicolor
[71] versicolor versicolor versicolor versicolor versicolor versicolor versicolor
[78] versicolor versicolor versicolor versicolor versicolor versicolor versicolor
[85] versicolor versicolor versicolor versicolor versicolor versicolor versicolor
[92] versicolor versicolor versicolor versicolor versicolor versicolor versicolor
[99] versicolor versicolor virginica virginica virginica virginica virginica
[106] virginica  virginica  virginica  virginica  virginica  virginica  virginica
[113] virginica  virginica  virginica  virginica  virginica  virginica  virginica
[120] virginica  virginica  virginica  virginica  virginica  virginica  virginica
[127] virginica  virginica  virginica  virginica  virginica  virginica  virginica
[134] virginica  virginica  virginica  virginica  virginica  virginica  virginica
[141] virginica  virginica  virginica  virginica  virginica  virginica  virginica
[148] virginica  virginica  virginica
Levels: setosa versicolor virginica

```

Filtering Data Frames

```

> set <- filterD(dfobj,species=="setosa")
> str(set)

```

```

'data.frame':  50 obs. of  5 variables:
 $ seplen : int  50 46 46 51 55 48 52 49 44 50 ...
 $ sepwid : int  33 34 36 33 35 31 34 36 32 35 ...
 $ petlen : int  14 14 10 17 13 16 14 14 13 16 ...
 $ petwid : int   2 3 2 5 2 2 2 1 2 6 ...
 $ species: Factor w/ 1 level "setosa": 1 1 1 1 1 1 1 1 1 1 ...

```

```

> gt5 <- filterD(dfobj,seplen>5)
> headtail(gt5)

```

| | seplen | sepwid | petlen | petwid | species |
|-----|--------|--------|--------|--------|-----------|
| 1 | 50 | 33 | 14 | 2 | setosa |
| 2 | 46 | 34 | 14 | 3 | setosa |
| 3 | 46 | 36 | 10 | 2 | setosa |
| 148 | 57 | 25 | 50 | 20 | virginica |
| 149 | 58 | 28 | 51 | 24 | virginica |
| 150 | 63 | 33 | 60 | 25 | virginica |

```

> setver1 <- filterD(dfobj,species!="virginica")
> str(setver1)

```

```

'data.frame':  100 obs. of  5 variables:
 $ seplen : int  50 46 46 51 55 48 52 49 44 50 ...
 $ sepwid : int  33 34 36 33 35 31 34 36 32 35 ...
 $ petlen : int  14 14 10 17 13 16 14 14 13 16 ...
 $ petwid : int   2 3 2 5 2 2 2 1 2 6 ...
 $ species: Factor w/ 2 levels "setosa","versicolor": 1 1 1 1 1 1 1 1 1 1 ...

```

```

> setver <- filterD(dfobj,species %in% c("setosa","versicolor"))
> str(setver)

```

```
'data.frame': 100 obs. of 5 variables:
 $ seplen : int 50 46 46 51 55 48 52 49 44 50 ...
 $ sepwid : int 33 34 36 33 35 31 34 36 32 35 ...
 $ petlen : int 14 14 10 17 13 16 14 14 13 16 ...
 $ petwid : int 2 3 2 5 2 2 2 1 2 6 ...
 $ species: Factor w/ 2 levels "setosa","versicolor": 1 1 1 1 1 1 1 1 1 1 ...
```

```
> ver_and_gt5 <- filterD(dfobj,species=="versicolor",seplen>5)
> headtail(ver_and_gt5)
```

| | seplen | sepwid | petlen | petwid | species |
|----|--------|--------|--------|--------|------------|
| 1 | 65 | 28 | 46 | 15 | versicolor |
| 2 | 62 | 22 | 45 | 15 | versicolor |
| 3 | 59 | 32 | 48 | 18 | versicolor |
| 48 | 66 | 30 | 44 | 14 | versicolor |
| 49 | 68 | 28 | 48 | 14 | versicolor |
| 50 | 67 | 30 | 50 | 17 | versicolor |

```
> ver_or_gt5 <- filterD(dfobj,species=="versicolor" | seplen>5)
> headtail(ver_or_gt5)
```

| | seplen | sepwid | petlen | petwid | species |
|-----|--------|--------|--------|--------|-----------|
| 1 | 50 | 33 | 14 | 2 | setosa |
| 2 | 46 | 34 | 14 | 3 | setosa |
| 3 | 46 | 36 | 10 | 2 | setosa |
| 148 | 57 | 25 | 50 | 20 | virginica |
| 149 | 58 | 28 | 51 | 24 | virginica |
| 150 | 63 | 33 | 60 | 25 | virginica |