

Filtering Data



Getting and Cleaning Data

```
> glimpse(msleep)
```

```
Observations: 83
```

```
Variables: 11
```

```
$ name      <chr> "Cheetah", "Owl monkey", "Mountain beaver", "Greater short-tailed shrew", "Cow", "Three...  
$ genus     <chr> "Acinonyx", "Aotus", "Aplodontia", "Blarina", "Bos", "Bradypus", "Callorhinus", "Calomy...  
$ vore      <chr> "carni", "omni", "herbi", "omni", "herbi", "herbi", "carni", NA, "carni", "herbi", "her...  
$ order     <chr> "Carnivora", "Primates", "Rodentia", "Soricomorpha", "Artiodactyla", "Pilosa", "Carnivo...  
$ conservation <chr> "lc", NA, "nt", "lc", "domesticated", NA, "vu", NA, "domesticated", "lc", "lc", "domest...  
$ sleep_total <dbl> 12.1, 17.0, 14.4, 14.9, 4.0, 14.4, 8.7, 7.0, 10.1, 3.0, 5.3, 9.4, 10.0, 12.5, 10.3, 8.3...  
$ sleep_rem  <dbl> NA, 1.8, 2.4, 2.3, 0.7, 2.2, 1.4, NA, 2.9, NA, 0.6, 0.8, 0.7, 1.5, 2.2, 2.0, 1.4, 3.1, ...  
$ sleep_cycle <dbl> NA, NA, NA, 0.133, 0.667, 0.767, 0.383, NA, 0.333, NA, NA, 0.217, NA, 0.117, NA, NA, 0....  
$ awake     <dbl> 11.9, 7.0, 9.6, 9.1, 20.0, 9.6, 15.3, 17.0, 13.9, 21.0, 18.7, 14.6, 14.0, 11.5, 13.7, 1...  
$ brainwt   <dbl> NA, 0.01550, NA, 0.00029, 0.42300, NA, NA, NA, 0.07000, 0.09820, 0.11500, 0.00550, NA, ...  
$ bodywt    <dbl> 50.000, 0.480, 1.350, 0.019, 600.000, 3.850, 20.490, 0.045, 14.000, 14.800, 33.500, 0.7...
```



> glimpse(msleep) There are:

Observations: 83

Variables: 11

- 83 rows
- 11 columns

```
$ name      <chr> "Cheetah", "Owl monkey", "Mountain beaver", "Greater short-tailed shrew", "Cow", "Three...
$ genus     <chr> "Acinonyx", "Aotus", "Aplodontia", "Blarina", "Bos", "Bradypus", "Callorhinus", "Calomy...
$vore       <chr> "carni", "omni", "herbi", "omni", "herbi", "herbi", "carni", NA, "carni", "herbi", "her...
$ order     <chr> "Carnivora", "Primates", "Rodentia", "Soricomorpha", "Artiodactyla", "Pilosa", "Carnivo...
$ conservation <chr> "lc", NA, "nt", "lc", "domesticated", NA, "vu", NA, "domesticated", "lc", "lc", "domest...
$ sleep_total <dbl> 12.1, 17.0, 14.4, 14.9, 4.0, 14.4, 8.7, 7.0, 10.1, 3.0, 5.3, 9.4, 10.0, 12.5, 10.3, 8.3...
$ sleep_rem  <dbl> NA, 1.8, 2.4, 2.3, 0.7, 2.2, 1.4, NA, 2.9, NA, 0.6, 0.8, 0.7, 1.5, 2.2, 2.0, 1.4, 3.1, ...
$ sleep_cycle <dbl> NA, NA, NA, 0.133, 0.667, 0.767, 0.383, NA, 0.333, NA, NA, 0.217, NA, 0.117, NA, NA, 0....
$ awake     <dbl> 11.9, 7.0, 9.6, 9.1, 20.0, 9.6, 15.3, 17.0, 13.9, 21.0, 18.7, 14.6, 14.0, 11.5, 13.7, 1...
$ brainwt    <dbl> NA, 0.01550, NA, 0.00029, 0.42300, NA, NA, NA, 0.07000, 0.09820, 0.11500, 0.00550, NA, ...
$ bodywt     <dbl> 50.000, 0.480, 1.350, 0.019, 600.000, 3.850, 20.490, 0.045, 14.000, 14.800, 33.500, 0.7...
```



> glimpse(msleep) There are:

Observations: 83

Variables: 11

- 83 rows
- 11 columns

```
$ name      <chr> "Cheetah", "Owl monkey", "Mountain beaver", "Greater short-tailed shrew", "Cow", "Three...
$ genus     <chr> "Acinonyx", "Aotus", "Aplodontia", "Blarina", "Bos", "Bradypus", "Callorhinus", "Calomy...
$vore       <chr> "carni", "omni", "herbi", "omni", "herbi", "herbi", "carni", NA, "carni", "herbi", "her...
$ order     <chr> "Carnivora", "Primates", "Rodentia", "Soricomorpha", "Artiodactyla", "Pilosa", "Carnivo...
$ conservation <chr> "lc", NA, "nt", "lc", "domesticated", NA, "vu", NA, "domesticated", "lc", "lc", "domest...
$ sleep_total <dbl> 12.1, 17.0, 14.4, 14.9, 4.0, 14.4, 8.7, 7.0, 10.1, 3.0, 5.3, 9.4, 10.0, 12.5, 10.3, 8.3...
$ sleep_rem  <dbl> NA, 1.8, 2.4, 2.3, 0.7, 2.2, 1.4, NA, 2.9, NA, 0.6, 0.8, 0.7, 1.5, 2.2, 2.0, 1.4, 3.1, ...
$ sleep_cycle <dbl> NA, NA, NA, 0.133, 0.667, 0.767, 0.383, NA, 0.333, NA, NA, 0.217, NA, 0.117, NA, NA, 0....
$ awake     <dbl> 11.9, 7.0, 9.6, 9.1, 20.0, 9.6, 15.3, 17.0, 13.9, 21.0, 18.7, 14.6, 14.0, 11.5, 13.7, 1...
$ brainwt    <dbl> NA, 0.01550, NA, 0.00029, 0.42300, NA, NA, NA, 0.07000, 0.09820, 0.11500, 0.00550, NA, ...
$ bodywt     <dbl> 50.000, 0.480, 1.350, 0.019, 600.000, 3.850, 20.490, 0.045, 14.000, 14.800, 33.500, 0.7...
```

The names of
the columns



```
> glimpse(msleep)
```

```
Observations: 83
```

```
Variables: 11
```

```
$ name      <chr> "Cheetah", "Owl monkey", "Mountain beaver", "Greater short-tailed shrew", "Cow", "Three...  
$ genus     <chr> "Acanonyx", "Aotus", "Aplodontia", "Blarina", "Bos", "Bradypus", "Callorhinus", "Calomy...  
$vore       <chr> "carni", "omni", "herbi", "omni", "herbi", "herbi", "carni", NA, "carni", "herbi", "her...  
$ order     <chr> "Carnivora", "Primates", "Rodentia", "Soricomorpha", "Artiodactyla", "Pilosa", "Carnivo...  
$ conservation <chr> "lc", NA, "nt", "lc", "domesticated", NA, "vu", NA, "domesticated", "lc", "lc", "domest...  
$ sleep_total <dbl> 12.1, 17.0, 14.4, 14.9, 4.0, 14.4, 8.7, 7.0, 10.1, 3.0, 5.3, 9.4, 10.0, 12.5, 10.3, 8.3...  
$ sleep_rem  <dbl> NA, 1.8, 2.4, 2.3, 0.7, 2.2, 1.4, NA, 2.9, NA, 0.6, 0.8, 0.7, 1.5, 2.2, 2.0, 1.4, 3.1, ...  
$ sleep_cycle <dbl> NA, NA, NA, 0.133, 0.667, 0.767, 0.383, NA, 0.333, NA, NA, 0.217, NA, 0.117, NA, NA, 0....  
$ awake     <dbl> 11.9, 7.0, 9.6, 9.1, 20.0, 9.6, 15.3, 17.0, 13.9, 21.0, 18.7, 14.6, 14.0, 11.5, 13.7, 1...  
$ brainwt   <dbl> NA, 0.01550, NA, 0.00029, 0.42300, NA, NA, NA, 0.07000, 0.09820, 0.11500, 0.00550, NA, ...  
$ bodywt    <dbl> 50.000, 0.480, 1.350, 0.019, 600.000, 3.850, 20.490, 0.045, 14.000, 14.800, 33.500, 0.7...
```

There are:

- 83 rows
- 11 columns

The first 5 columns are
character variables

The names of
the columns




```
> glimpse(msleep)
```

```
Observations: 83
```

```
Variables: 11
```

```
$ name      <chr> "Cheetah", "Owl monkey", "Mountain beaver", "Greater short-tailed shrew", "Cow", "Three...  
$ genus     <chr> "Acinonyx", "Aotus", "Aplodontia", "Blarina", "Bos", "Bradypus", "Callorhinus", "Calomy...  
$vore       <chr> "carni", "omni", "herbi", "omni", "herbi", "herbi", "carni", NA, "carni", "herbi", "her...  
$ order     <chr> "Carnivora", "Primates", "Rodentia", "Soricomorpha", "Artiodactyla", "Pilosa", "Carnivo...  
$ conservation <chr> "lc", NA, "nt", "lc", "domesticated", NA, "vu", NA, "domesticated", "lc", "lc", "domest...  
$ sleep_total <dbl> 12.1, 17.0, 14.4, 14.9, 4.0, 14.4, 8.7, 7.0, 10.1, 3.0, 5.3, 9.4, 10.0, 12.5, 10.3, 8.3...  
$ sleep_rem  <dbl> NA, 1.8, 2.4, 2.3, 0.7, 2.2, 1.4, NA, 2.9, NA, 0.6, 0.8, 0.7, 1.5, 2.2, 2.0, 1.4, 3.1, ...  
$ sleep_cycle <dbl> NA, NA, NA, 0.133, 0.667, 0.767, 0.383, NA, 0.333, NA, NA, 0.217, NA, 0.117, NA, NA, 0....  
$ awake     <dbl> 11.9, 7.0, 9.6, 9.1, 20.0, 9.6, 15.3, 17.0, 13.9, 21.0, 18.7, 14.6, 14.0, 11.5, 13.7, 1...  
$ brainwt   <dbl> NA, 0.01550, NA, 0.00029, 0.42300, NA, NA, NA, 0.07000, 0.09820, 0.11500, 0.00550, NA, ...  
$ bodywt    <dbl> 50.000, 0.480, 1.350, 0.019, 600.000, 3.850, 20.490, 0.045, 14.000, 14.800, 33.500, 0.7...
```

There are:

- 83 rows
- 11 columns

The first 5 columns are
character variables

The first three names of the
animals in the dataset

The names of
the columns



Equivalent to:

```
> msleep %>%  
+   filter(order == "Primates")
```

```
filter(msleep, order == "Primates")
```

```
# A tibble: 12 x 11
```

	name	genus	vore	order	conservation	sleep_total	sleep_rem	sleep_cycle	awake	brainwt	bodywt
	<chr>	<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	Owl monkey	Aotus	omni	Prima...	<NA>	17.0	1.80	NA	7.00	0.0155	0.480
2	Griquet	Cercopithe...	omni	Prima...	lc	10.0	0.700	NA	14.0	NA	4.75
3	Patas monkey	Erythroceb...	omni	Prima...	lc	10.9	1.10	NA	13.1	0.115	10.0
4	Galago	Galago	omni	Prima...	<NA>	9.80	1.10	0.550	14.2	0.00500	0.200
5	Human	Homo	omni	Prima...	<NA>	8.00	1.90	1.50	16.0	1.32	62.0
6	Mongoose lemur	Lemur	herbi	Prima...	vu	9.50	0.900	NA	14.5	NA	1.67
7	Macaque	Macaca	omni	Prima...	<NA>	10.1	1.20	0.750	13.9	0.179	6.80
8	Slow loris	Nyctibeus	carni	Prima...	<NA>	11.0	NA	NA	13.0	0.0125	1.40
9	Chimpanzee	Pan	omni	Prima...	<NA>	9.70	1.40	1.42	14.3	0.440	52.2
10	Baboon	Papio	omni	Prima...	<NA>	9.40	1.00	0.667	14.6	0.180	25.2
11	Potto	Perodictic...	omni	Prima...	lc	11.0	NA	NA	13.0	NA	1.10
12	Squirrel monkey	Saimiri	omni	Prima...	<NA>	9.60	1.40	NA	14.4	0.0200	0.743



```
# Filter using base R and the square bracket operator
```

```
msleep[msleep$order == "Primates", ]
```

```
# A tibble: 12 x 11
```

	name	genus	vore	order	conservation	sleep_total	sleep_rem
	<chr>	<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>
1	Owl ...	Aotus	omni	Prim...	NA	17	1.8
2	Griv...	Cerc...	omni	Prim...	lc	10	0.7
3	Pata...	Eryt...	omni	Prim...	lc	10.9	1.1
4	Gala...	Gala...	omni	Prim...	NA	9.8	1.1
5	Human	Homo	omni	Prim...	NA	8	1.9
6	Mong...	Lemur	herbi	Prim...	vu	9.5	0.9
7	Maca...	Maca...	omni	Prim...	NA	10.1	1.2
8	Slow...	Nyct...	carni	Prim...	NA	11	NA
9	Chim...	Pan	omni	Prim...	NA	9.7	1.4
10	Babo...	Papio	omni	Prim...	NA	9.4	1
11	Potto	Pero...	omni	Prim...	lc	11	NA
12	Squi...	Saim...	omni	Prim...	NA	9.6	1.4

```
# ... with 4 more variables: sleep_cycle <dbl>, awake <dbl>,
```

```
# brainwt <dbl>, bodywt <dbl>
```



```
# Save the result of our filtering to an object
```

```
sleepy_primates <- msleep %>%  
  filter(order == "Primates")
```

```
# Plot each species' amount of sleep
```

```
ggplot(sleepy_primates,      # plot data from the sleepy_primates object  
  aes(x=name, y=sleep_total)) + # with the species name on the x and the amount of sleep on the y axis  
  geom_bar(stat="identity") +  # use a bar plot  
  coord_flip()                 # flip the x and y axes so that the species names are horizontal and easy to read
```



name

Squirrel monkey

Slow loris

Potto

Patas monkey

Owl monkey

Mongoose lemur

Macaque

Human

Grivet

Galago

Chimpanzee

Baboon

0

5

10

15

sleep_total



```
> msleep %>%  
+   filter(order == "Primates", sleep_total > 10)
```

```
# A tibble: 5 x 11
```

	name	genus	vore	order	conservation	sleep_total	sleep_rem	sleep_cycle	awake	brainwt	bodywt
	<chr>	<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	Owl monkey	Aotus	omni	Primates	<NA>	17.0	1.80	NA	7.00	0.0155	0.480
2	Patas monkey	Erythrocebus	omni	Primates	lc	10.9	1.10	NA	13.1	0.115	10.0
3	Macaque	Macaca	omni	Primates	<NA>	10.1	1.20	0.750	13.9	0.179	6.80
4	Slow loris	Nyctibeus	carni	Primates	<NA>	11.0	NA	NA	13.0	0.0125	1.40
5	Potto	Perodicticus	omni	Primates	lc	11.0	NA	NA	13.0	NA	1.10

Gives the same results: `msleep %>%
 filter(order == "Primates" & sleep_total > 10)`



```
> msleep %>%  
+   filter(order == "Primates", sleep_total > 10) %>%  
+   select(name, sleep_total, sleep_rem, sleep_cycle)  
# A tibble: 5 x 4
```

	name	sleep_total	sleep_rem	sleep_cycle
	<chr>	<dbl>	<dbl>	<dbl>
1	Owl monkey	17.0	1.80	NA
2	Patas monkey	10.9	1.10	NA
3	Macaque	10.1	1.20	0.750
4	Slow loris	11.0	NA	NA
5	Potto	11.0	NA	NA

Without the pipe operator: `select(filter(msleep, order == "Primates",
sleep_total > 10), name, sleep_total, sleep_rem,
sleep_cycle)`



```
> msleep %>%  
+   filter(order == "Primates", sleep_total > 10) %>%  
+   select(name, total=sleep_total, rem=sleep_rem, cycle=sleep_cycle)
```

```
# A tibble: 5 x 4
```

	name	total	rem	cycle
	<chr>	<dbl>	<dbl>	<dbl>
1	Owl monkey	17.0	1.80	NA
2	Patas monkey	10.9	1.10	NA
3	Macaque	10.1	1.20	0.750
4	Slow loris	11.0	NA	NA
5	Potto	11.0	NA	NA




```

> msleep %>%
+   filter(order == "Primates", sleep_total > 10) %>%
+   rename(total=sleep_total, rem=sleep_rem, cycle=sleep_cycle)
# A tibble: 5 x 11
  name      genus      vore order conservation total rem cycle awake brainwt bodywt
  <chr>    <chr>    <chr> <chr>    <chr>    <dbl> <dbl> <dbl> <dbl>    <dbl> <dbl>
1 Owl monkey Aotus      omni  Primates <NA>    17    1.8 NA     7    0.0155  0.48
2 Patas monkey Erythrocebus omni  Primates lc     10.9    1.1 NA    13.1  0.115  10
3 Macaque     Macaca      omni  Primates <NA>    10.1    1.2  0.75  13.9  0.179   6.8
4 Slow loris  Nyctibeus   carni  Primates <NA>    11     NA   NA    13    0.0125  1.4
5 Potto      Perodicticus omni  Primates lc     11     NA   NA    13    NA     1.1

```



Summarizing: Filtering Data



Getting and Cleaning Data