

Summarizing Data



Getting and Cleaning Data

```
> msleep
# A tibble: 83 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 Cheetah  Acinonyx  carni Carnivo... lc          12.1      NA          NA     11.9  NA      5.00e+1
2 Owl monkey Aotus    omni  Primates <NA>        17.0      1.80      NA      7.00  1.55e-2  4.80e-1
3 Mountain beaver Aplodon... herbi Rodentia nt         14.4      2.40      NA      9.60  NA      1.35e+0
4 Greater short-... Blarina  omni  Soricom... lc          14.9      2.30      0.133  9.10  2.90e-4  1.90e-2
5 Cow       Bos      herbi Artioda... domesticated 4.00      0.700      0.667  20.0  4.23e-1  6.00e+2
6 Three-toed slo... Bradypus herbi Pilosa  <NA>        14.4      2.20      0.767  9.60  NA      3.85e+0
7 Northern fur s... Callorh... carni Carnivo... vu          8.70      1.40      0.383  15.3  NA      2.05e+1
8 Vesper mouse Calomys  <NA>  Rodentia <NA>        7.00      NA          NA     17.0  NA      4.50e-2
9 Dog       Canis    carni Carnivo... domesticated 10.1      2.90      0.333  13.9  7.00e-2  1.40e+1
10 Roe deer  Capreol... herbi Artioda... lc          3.00      NA          NA     21.0  9.82e-2  1.48e+1
# ... with 73 more rows
```

```
> msleep %>%
+   group_by(order)
# A tibble: 83 x 11
# Groups:   order [19]
  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 Cheetah  Acinonyx  carni Carnivo... lc          12.1      NA          NA     11.9  NA      5.00e+1
2 Owl monkey Aotus    omni  Primates <NA>        17.0      1.80      NA      7.00  1.55e-2  4.80e-1
3 Mountain beaver Aplodon... herbi Rodentia nt         14.4      2.40      NA      9.60  NA      1.35e+0
4 Greater short-... Blarina  omni  Soricom... lc          14.9      2.30      0.133  9.10  2.90e-4  1.90e-2
5 Cow       Bos      herbi Artioda... domesticated 4.00      0.700      0.667  20.0  4.23e-1  6.00e+2
6 Three-toed slo... Bradypus herbi Pilosa  <NA>        14.4      2.20      0.767  9.60  NA      3.85e+0
7 Northern fur s... Callorh... carni Carnivo... vu          8.70      1.40      0.383  15.3  NA      2.05e+1
8 Vesper mouse Calomys  <NA>  Rodentia <NA>        7.00      NA          NA     17.0  NA      4.50e-2
9 Dog       Canis    carni Carnivo... domesticated 10.1      2.90      0.333  13.9  7.00e-2  1.40e+1
10 Roe deer  Capreol... herbi Artioda... lc          3.00      NA          NA     21.0  9.82e-2  1.48e+1
# ... with 73 more rows
```

```
msleep %>%
```

```
summarize(N=n())
```

```
# A tibble: 1 x 1
```

```
      N
```

```
  <int>
```

```
1     83
```

```
.
```

← New column name, N

← Same as `nrow(msleep)`



```
> msleep %>%  
+ group_by(order) %>%  
+ summarize(N=n())  
# A tibble: 19 x 2
```

	order <chr>	N <int>
1	Afrosoricida	1
2	Artiodactyla	6
3	Carnivora	12
4	Cetacea	3
5	Chiroptera	2
6	Cingulata	2
7	Didelphimorphia	2
8	Diprotodontia	2
9	Erinaceomorpha	2
10	Hyracoidea	3
11	Lagomorpha	1
12	Monotremata	1
13	Perissodactyla	3
14	Pilosa	1
15	Primates	12
16	Proboscidea	2
17	Rodentia	22
18	Scandentia	1
19	Soricomorpha	5



```
> msleep %>%  
+   group_by(order) %>%  
+   summarize(N=n(), mean_sleep=mean(sleep_total))
```

```
# A tibble: 19 x 3
```

	order <chr>	N <int>	mean_sleep <dbl>
1	Afrosoricida	1	15.6
2	Artiodactyla	6	4.52
3	Carnivora	12	10.1
4	Cetacea	3	4.5
5	Chiroptera	2	19.8
6	Cingulata	2	17.8
7	Didelphimorphia	2	18.7
8	Diprotodontia	2	12.4
9	Erinaceomorpha	2	10.2
10	Hyracoidea	3	5.67
11	Lagomorpha	1	8.4
12	Monotremata	1	8.6
13	Perissodactyla	3	3.47
14	Pilosa	1	14.4
15	Primates	12	10.5
16	Proboscidea	2	3.6
17	Rodentia	22	12.5
18	Scandentia	1	8.9
19	Soricomorpha	5	11.1

```
> msleep %>%
```

```
+ tabyl(order)
```

		order	n	percent
1		Afrosoricida	1	0.0120
2		Artiodactyla	6	0.0723
3		Carnivora	12	0.1446
4		Cetacea	3	0.0361
5		Chiroptera	2	0.0241
6		Cingulata	2	0.0241
7		Didelphimorphia	2	0.0241
8		Diprotodontia	2	0.0241
9		Erinaceomorpha	2	0.0241
10		Hyracoidea	3	0.0361
11		Lagomorpha	1	0.0120
12		Monotremata	1	0.0120
13		Perissodactyla	3	0.0361
14		Pilosa	1	0.0120
15		Primates	12	0.1446
16		Proboscidea	2	0.0241
17		Rodentia	22	0.2651
18		Scandentia	1	0.0120
19		Soricomorpha	5	0.0602



```
summary(msleep$sleep_total)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.90	7.85	10.10	10.43	13.75	19.90

summary() is appropriate
for summarizing
numeric variables

When you apply tabyl() to a numerical variable, it still
treats it as a categorical variable and will count the
number of times that number appears

```
tabyl(msleep$sleep_total)
```

msleep\$sleep_total	n	percent
1.9	1	0.01204819
2.7	1	0.01204819
2.9	1	0.01204819
3.0	1	0.01204819
3.1	1	0.01204819
3.3	1	0.01204819
3.5	1	0.01204819
3.8	1	0.01204819
3.9	1	0.01204819
4.0	1	0.01204819
4.4	1	0.01204819
5.2	1	0.01204819
5.3	2	0.02409639
5.4	1	0.01204819
5.6	1	0.01204819
6.2	1	0.01204819
6.3	2	0.02409639
7.0	1	0.01204819
7.7	1	0.01204819
8.0	1	0.01204819



```
> skim(msleep)
```

Skim summary statistics







n obs: 83

n variables: 11

Variable type: character

variable	missing	complete	n	min	max	empty	n_unique
conservation	29	54	83	2	12	0	6
genus	0	83	83	3	13	0	77
name	0	83	83	3	30	0	83
order	0	83	83	6	15	0	19
vore	7	76	83	4	7	0	4

Variable type: numeric

variable	missing	complete	n	mean	sd	p0	p25	p50	p75	p100	hist
awake	0	83	83	13.57	4.45	4.1	10.25	13.9	16.15	22.1	
bodywt	0	83	83	166.14	786.84	0.005	0.17	1.67	41.75	6654	
brainwt	27	56	83	0.28	0.98	0.00014	0.0029	0.012	0.13	5.71	
sleep_cycle	51	32	83	0.44	0.36	0.12	0.18	0.33	0.58	1.5	
sleep_rem	22	61	83	1.88	1.3	0.1	0.9	1.5	2.4	6.6	
sleep_total	0	83	83	10.43	4.45	1.9	7.85	10.1	13.75	19.9	



Summarizing: Summarizing Data



Getting and Cleaning Data