Manipulating columns

Getting and Cleaning Data

```
msleep %>%
  filter(order == "Primates", sleep_total > 10) %>%
  select(name, sleep_rem, sleep_cycle, sleep_total) %>%
  arrange(name) %>%
  mutate(sleep_total_min = sleep_total * 60)
```

A tibble: 5 x 5

name	sleep_rem	sleep_cycle	sleep_total	sleep_total_min
<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1 Macaque	1.2	0.75	10.1	606
2 Owl monkey	1.8	NA	17	1020
3 Patas monkey	1.1	NA	10.9	654
4 Potto	NA	NA	11	660
5 Slow loris	NA	NA	11	660

A whole new column!

```
msleep %>%
  filter(order == "Primates", sleep_total > 10) %>%
  select(name, sleep_rem, sleep_cycle, sleep_total) %>%
  arrange(name) %>%
  mutate(new_sleep = "msleep_data")
```

A tibble: 5 x 5

name	sleep_rem	sleep_cycle	sleep_total	new_sleep
<chr>	<dbl>	<dbl>	<dbl>	<chr>
1 Macaque	1.2	0.75	10.1	msleep_data
2 Owl monkey	1.8	NA	17	msleep_data
3 Patas monkey	1.1	NA	10.9	msleep_data
4 Potto	NA	NA	11	msleep_data
5 Slow loris	NA	NA	11	msleep_data

A whole new column that is equal to the same value in all rows

```r
# Save the results of filter, select and arrange to a dataframe
sleepy_primates <- msleep %>%
  filter(order == "Primates", sleep_total > 10) %>%
  select(name, sleep_rem, sleep_cycle, sleep_total) %>%
  arrange(name)

# To create a new column with the sleep total converted to minutes
sleepy_primates$sleep_total_min <- sleepy_primates$sleep_total * 60

# To add a column with the same value in each row
sleepy_primates$new_sleep <- "msleep_data"
```

```
# Getting rid of the column bodywt
msleep %>%
  filter(order == "Primates", sleep_total > 10) %>%
  select(-bodywt)
```

Now our dataset has only 10 columns instead of 11!

A tibble: 5 x 10

name	genus	vore	order	conservation	sleep_total	sleep_rem	sleep_cycle	awake	brainwt
<chr>	<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1 Owl monkey	Aotus	omni	Primates	NA	17	1.8	NA	7	0.0155
2 Patas monkey	Erythrocebus	omni	Primates	lc	10.9	1.1	NA	13.1	0.115
3 Macaque	Macaca	omni	Primates	NA	10.1	1.2	0.75	13.9	0.179
4 Slow loris	Nyctibeus	carni	Primates	NA	11	NA	NA	13	0.0125
5 Potto	Perodicticus	omni	Primates	lc	11	NA	NA	13	NA

```
> conservation
# A tibble: 11 x 1
   `conservation abbreviation`
   <chr>
 1 EX = Extinct
 2 EW = Extinct in the wild
 3 CR = Critically Endangered
 4 EN = Endangered
 5 VU = Vulnerable
 6 NT = Near Threatened
 7 LC = Least Concern
 8 DD = Data deficient
 9 NE = Not evaluated
10 PE = Probably extinct (informal)
11 PEW = Probably extinct in the wild (informal)
```

Tidy data violation!
Space in column name should be an underscore.

Tidy data violation!
There are two pieces of information in a column: (1) abbreviation and (2) description.

```
> conservation %>%
+    separate(`conservation abbreviation`,
+             into = c("abbreviation", "description"), sep = " = ")
# A tibble: 11 x 2
   abbreviation description
   <chr>        <chr>
 1 EX           Extinct
 2 EW           Extinct in the wild
 3 CR           Critically Endangered
 4 EN           Endangered
 5 VU           Vulnerable
 6 NT           Near Threatened
 7 LC           Least Concern
 8 DD           Data deficient
 9 NE           Not evaluated
10 PE           Probably extinct (informal)
11 PEW          Probably extinct in the wild (informal)
```

```
> conservation %>%
+   separate(`conservation abbreviation`,
+           into = c("abbreviation", "description"), sep = " = ") %>%
+   unite(united_col, abbreviation, description, sep=" = ")
# A tibble: 11 x 1
   united_col
   <chr>
 1 EX = Extinct
 2 EW = Extinct in the wild
 3 CR = Critically Endangered
 4 EN = Endangered
 5 VU = Vulnerable
 6 NT = Near Threatened
 7 LC = Least Concern
 8 DD = Data deficient
 9 NE = Not evaluated
10 PE = Probably extinct (informal)
11 PEW = Probably extinct in the wild (informal)
```

```
> conservation %>%
+   clean_names()
# A tibble: 11 x 1
   conservation_abbreviation
   <chr>
 1 EX = Extinct
 2 EW = Extinct in the wild
 3 CR = Critically Endangered
 4 EN = Endangered
 5 VU = Vulnerable
 6 NT = Near Threatened
 7 LC = Least Concern
 8 DD = Data deficient
 9 NE = Not evaluated
10 PE = Probably extinct (informal)
11 PEW = Probably extinct in the wild (informal)
```

Adds underscore to column name

Summarizing: Manipulating Data

Getting and Cleaning Data