

## DPLYR CHEAT SHEET

#important verbs to remember

# dplyr verbs	Description
# select()	select columns
# filter()	filter rows
# arrange()	re-order or arrange rows
# mutate()	create new columns
# summarise()	summarise values
# group_by()	allows for group operations in the “split-apply-combine” concept

```
mammal_data =  
read.csv("https://raw.githubusercontent.com/genomicsclass/dagdata/master/inst/extdata/msleep_ggplot2.csv",st  
ringsAsFactors = FALSE)
```

```
library(dplyr)
```

```
fix(mammal_data)
```

#select specific columns

```
sleep_data = select(mammal_data,name,sleep_total)
```

#select all columns except specific column

```
sleep_data_without_name = select(sleep_data,-name)
```

#selecting a range of columns

```
range_of_columns = select(mammal_data,name:order)
```

# Some additional options to select columns based on a specific criteria include

#

# ends\_with() = Select columns that end with a character string

# contains() = Select columns that contain a character string

# matches() = Select columns that match a regular expression

# one\_of() = Select columns names that are from a group of names

```
columns_starting_with_sleep = select(mammal_data,starts_with("sleep"))
```

```
columns_ending_with_vore = select(mammal_data,ends_with("vore"))
```

```
columns_containing_wt = select(mammal_data,contains("wt"))
```

```
columns_containing_one_of = select(mammal_data,one_of("height","age","name"))
```

#Rows can be filtered using filter()

```
filter(sleep_data, sleep_total>15 )
```

```
filter(mammal_data, sleep_total>15 , bodywt > 1)
```

```
filter(mammal_data, order %in% c("Carnivora","Rodentia"), conservation == "domesticated")
```

# pipe operator %>%

# head(select(mammal\_data,name,sleep\_total)) can also be written like this

```
mammal_data %>% select(name,sleep_total) %>% head
```

```
mammal_data %>% arrange(order) %>% head
```

# Now, we will select three columns from msleep,

# arrange the rows by the taxonomic order and

# then arrange the rows by sleep\_total.

# Finally show the head of the final data frame

```
mammal_data %>% select(name,order,sleep_total) %>% arrange(order,sleep_total) %>% head
```

# Same as above, except here we filter the rows for mammals that sleep for 16

# or more hours instead of showing the head of the final data frame

```
mammal_data %>% select(name,order,sleep_total) %>% arrange(order,sleep_total) %>% filter(sleep_total > 16)
```

# Something slightly more complicated: same as above, except arrange the rows in the sleep\_total

# column in a descending order. For this, use the function desc()

```
mammal_data %>% select(name,order,sleep_total) %>% arrange(order,desc(sleep_total)) %>% filter(sleep_total > 16)
```

#add new columns using mutate

```
mammal_data %>% mutate(rem_proportion = (sleep_rem/sleep_total)*100, body_wt_grams = bodywt*1000) %>% head
```

#summarise can be applied to columns to get mean , median etc.,

```
mammal_data %>% summarise(avg_sleep = mean(sleep_total))
```

```
mammal_data %>% summarise(avg_sleep = mean(sleep_total),min_sleep = min(sleep_total),max_sleep = max(sleep_total),total = n())
```

#Group By: split the msleep data frame by the taxonomic order, then ask for the same summary statistics as above.

#We expect a set of summary statistics for each taxonomic order.

```
mammal_data %>% group_by(order) %>% summarise(avg_sleep = mean(sleep_total),min_sleep = min(sleep_total), max_sleep = max(sleep_total),total = n())
```