R Notebook

This is an [R Markdown](http://rmarkdown.rstudio.com) Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Cmd+Shift+Enter*.

plot(cars)



Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Cmd+Option+I*.

#Manipulate your data. Data wrangling. R programmning with Greg Martin

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.4 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

# loads the package  
library(dplyr)   
  
?msleep # to know more about the msleep data set   
glimpse(msleep)

## Rows: 83  
## Columns: 11  
## $ name <chr> "Cheetah", "Owl monkey", "Mountain beaver", "Greater shor…  
## $ genus <chr> "Acinonyx", "Aotus", "Aplodontia", "Blarina", "Bos", "Bra…  
## $ vore <chr> "carni", "omni", "herbi", "omni", "herbi", "herbi", "carn…  
## $ order <chr> "Carnivora", "Primates", "Rodentia", "Soricomorpha", "Art…  
## $ conservation <chr> "lc", NA, "nt", "lc", "domesticated", NA, "vu", NA, "dome…  
## $ sleep\_total <dbl> 12.1, 17.0, 14.4, 14.9, 4.0, 14.4, 8.7, 7.0, 10.1, 3.0, 5…  
## $ sleep\_rem <dbl> NA, 1.8, 2.4, 2.3, 0.7, 2.2, 1.4, NA, 2.9, NA, 0.6, 0.8, …  
## $ sleep\_cycle <dbl> NA, NA, NA, 0.1333333, 0.6666667, 0.7666667, 0.3833333, N…  
## $ awake <dbl> 11.9, 7.0, 9.6, 9.1, 20.0, 9.6, 15.3, 17.0, 13.9, 21.0, 1…  
## $ brainwt <dbl> NA, 0.01550, NA, 0.00029, 0.42300, NA, NA, NA, 0.07000, 0…  
## $ bodywt <dbl> 50.000, 0.480, 1.350, 0.019, 600.000, 3.850, 20.490, 0.04…

# Variable is renamed  
msleep %>%   
 rename("conserv" = "conservation")

## # A tibble: 83 × 11  
## name genus vore order conserv sleep\_total sleep\_rem sleep\_cycle awake  
## <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 Cheetah Acin… carni Carn… lc 12.1 NA NA 11.9  
## 2 Owl monkey Aotus omni Prim… <NA> 17 1.8 NA 7   
## 3 Mountain b… Aplo… herbi Rode… nt 14.4 2.4 NA 9.6  
## 4 Greater sh… Blar… omni Sori… lc 14.9 2.3 0.133 9.1  
## 5 Cow Bos herbi Arti… domest… 4 0.7 0.667 20   
## 6 Three-toed… Brad… herbi Pilo… <NA> 14.4 2.2 0.767 9.6  
## 7 Northern f… Call… carni Carn… vu 8.7 1.4 0.383 15.3  
## 8 Vesper mou… Calo… <NA> Rode… <NA> 7 NA NA 17   
## 9 Dog Canis carni Carn… domest… 10.1 2.9 0.333 13.9  
## 10 Roe deer Capr… herbi Arti… lc 3 NA NA 21   
## # ℹ 73 more rows  
## # ℹ 2 more variables: brainwt <dbl>, bodywt <dbl>

# rename() function is to rename names of variables of rows or columns  
head(msleep)

## # A tibble: 6 × 11  
## name genus vore order conservation sleep\_total sleep\_rem sleep\_cycle awake  
## <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 Cheetah Acin… carni Carn… lc 12.1 NA NA 11.9  
## 2 Owl mo… Aotus omni Prim… <NA> 17 1.8 NA 7   
## 3 Mounta… Aplo… herbi Rode… nt 14.4 2.4 NA 9.6  
## 4 Greate… Blar… omni Sori… lc 14.9 2.3 0.133 9.1  
## 5 Cow Bos herbi Arti… domesticated 4 0.7 0.667 20   
## 6 Three-… Brad… herbi Pilo… <NA> 14.4 2.2 0.767 9.6  
## # ℹ 2 more variables: brainwt <dbl>, bodywt <dbl>

# prints first few rows of the renamed table

# Reorder variables  
msleep %>%   
 select(vore,name,everything())

## # A tibble: 83 × 11  
## vore name genus order conservation sleep\_total sleep\_rem sleep\_cycle awake  
## <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 carni Cheet… Acin… Carn… lc 12.1 NA NA 11.9  
## 2 omni Owl m… Aotus Prim… <NA> 17 1.8 NA 7   
## 3 herbi Mount… Aplo… Rode… nt 14.4 2.4 NA 9.6  
## 4 omni Great… Blar… Sori… lc 14.9 2.3 0.133 9.1  
## 5 herbi Cow Bos Arti… domesticated 4 0.7 0.667 20   
## 6 herbi Three… Brad… Pilo… <NA> 14.4 2.2 0.767 9.6  
## 7 carni North… Call… Carn… vu 8.7 1.4 0.383 15.3  
## 8 <NA> Vespe… Calo… Rode… <NA> 7 NA NA 17   
## 9 carni Dog Canis Carn… domesticated 10.1 2.9 0.333 13.9  
## 10 herbi Roe d… Capr… Arti… lc 3 NA NA 21   
## # ℹ 73 more rows  
## # ℹ 2 more variables: brainwt <dbl>, bodywt <dbl>

# Selects the columns 'vore', 'name'  
#everything is used to select all other columns  
  
#chnaging varaible type  
class(msleep$vore)

## [1] "character"

# shows the class of vore varaible in dataser msleep  
msleep$vore <- as.factor(msleep$vore)   
# converting it into a factor  
  
# there is another way to change the varaible name   
msleep %>%  
 mutate(vore = as.character(vore)) %>%   
 # again chnaging variable to character  
 head() # prints first few rows of updated data set

## # A tibble: 6 × 11  
## name genus vore order conservation sleep\_total sleep\_rem sleep\_cycle awake  
## <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 Cheetah Acin… carni Carn… lc 12.1 NA NA 11.9  
## 2 Owl mo… Aotus omni Prim… <NA> 17 1.8 NA 7   
## 3 Mounta… Aplo… herbi Rode… nt 14.4 2.4 NA 9.6  
## 4 Greate… Blar… omni Sori… lc 14.9 2.3 0.133 9.1  
## 5 Cow Bos herbi Arti… domesticated 4 0.7 0.667 20   
## 6 Three-… Brad… herbi Pilo… <NA> 14.4 2.2 0.767 9.6  
## # ℹ 2 more variables: brainwt <dbl>, bodywt <dbl>

#selecting variables to work with  
names(msleep)

## [1] "name" "genus" "vore" "order" "conservation"  
## [6] "sleep\_total" "sleep\_rem" "sleep\_cycle" "awake" "brainwt"   
## [11] "bodywt"

# prints the names of all columns of msleep dataset  
 msleep %>%   
 select(2:4,   
 # selects 2,3,4 varaibles   
 awake, # adds awake  
 starts\_with("sleep"),   
 # selects all columns start with sleep  
 contains ("wt")) %>%   
 # selects the colums with wt in their varaible name   
 names()

## [1] "genus" "vore" "order" "awake" "sleep\_total"  
## [6] "sleep\_rem" "sleep\_cycle" "brainwt" "bodywt"

# prints the names of the selected

#filter and arrange data  
unique(msleep$order)

## [1] "Carnivora" "Primates" "Rodentia" "Soricomorpha"   
## [5] "Artiodactyla" "Pilosa" "Cingulata" "Hyracoidea"   
## [9] "Didelphimorphia" "Proboscidea" "Chiroptera" "Perissodactyla"   
## [13] "Erinaceomorpha" "Cetacea" "Lagomorpha" "Diprotodontia"   
## [17] "Monotremata" "Afrosoricida" "Scandentia"

#Returns the unique values from the order column

msleep %>%   
 filter((order== "Carnivora" |   
 order== "Primates")&  
 # Filter rows where the order is either "Carnivora" or "Primates"  
 sleep\_total > 8) %>%   
 # and total sleep duration is greater than 8 hours  
 select(name,order,sleep\_total) %>%   
 # Select only the columns: name, order, and total sleep   
 arrange(-sleep\_total)

## # A tibble: 20 × 3  
## name order sleep\_total  
## <chr> <chr> <dbl>  
## 1 Owl monkey Primates 17   
## 2 Tiger Carnivora 15.8  
## 3 Lion Carnivora 13.5  
## 4 Domestic cat Carnivora 12.5  
## 5 Arctic fox Carnivora 12.5  
## 6 Cheetah Carnivora 12.1  
## 7 Slow loris Primates 11   
## 8 Potto Primates 11   
## 9 Patas monkey Primates 10.9  
## 10 Jaguar Carnivora 10.4  
## 11 Dog Carnivora 10.1  
## 12 Macaque Primates 10.1  
## 13 Grivet Primates 10   
## 14 Galago Primates 9.8  
## 15 Red fox Carnivora 9.8  
## 16 Chimpanzee Primates 9.7  
## 17 Squirrel monkey Primates 9.6  
## 18 Mongoose lemur Primates 9.5  
## 19 Baboon Primates 9.4  
## 20 Northern fur seal Carnivora 8.7

#this will Arrange the data in descending order of total sleep duration

#this ia another way fo above shown method  
data(msleep)  
print(head(msleep))

## # A tibble: 6 × 11  
## name genus vore order conservation sleep\_total sleep\_rem sleep\_cycle awake  
## <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 Cheetah Acin… carni Carn… lc 12.1 NA NA 11.9  
## 2 Owl mo… Aotus omni Prim… <NA> 17 1.8 NA 7   
## 3 Mounta… Aplo… herbi Rode… nt 14.4 2.4 NA 9.6  
## 4 Greate… Blar… omni Sori… lc 14.9 2.3 0.133 9.1  
## 5 Cow Bos herbi Arti… domesticated 4 0.7 0.667 20   
## 6 Three-… Brad… herbi Pilo… <NA> 14.4 2.2 0.767 9.6  
## # ℹ 2 more variables: brainwt <dbl>, bodywt <dbl>

msleep %>%   
 # includes the data set and piping to next function  
 filter(order %in% c("Carnivora", "Primates")   
 & sleep\_total > 8) %>%   
 select(name,order,sleep\_total) %>%   
 # Filters rows for 'Carnivora' or 'Primates'   
 #and 'sleep\_total' more than 8  
 select(name,order,sleep\_total) %>%   
 arrange(order) # arranges data by the order

## # A tibble: 20 × 3  
## name order sleep\_total  
## <chr> <chr> <dbl>  
## 1 Cheetah Carnivora 12.1  
## 2 Northern fur seal Carnivora 8.7  
## 3 Dog Carnivora 10.1  
## 4 Domestic cat Carnivora 12.5  
## 5 Tiger Carnivora 15.8  
## 6 Jaguar Carnivora 10.4  
## 7 Lion Carnivora 13.5  
## 8 Arctic fox Carnivora 12.5  
## 9 Red fox Carnivora 9.8  
## 10 Owl monkey Primates 17   
## 11 Grivet Primates 10   
## 12 Patas monkey Primates 10.9  
## 13 Galago Primates 9.8  
## 14 Mongoose lemur Primates 9.5  
## 15 Macaque Primates 10.1  
## 16 Slow loris Primates 11   
## 17 Chimpanzee Primates 9.7  
## 18 Baboon Primates 9.4  
## 19 Potto Primates 11   
## 20 Squirrel monkey Primates 9.6

library(tidyverse)   
# loads the package  
library(dplyr)   
#change obseravations  
msleep %>%   
 mutate(brainwt= brainwt\*1000)

## # A tibble: 83 × 11  
## name genus vore order conservation sleep\_total sleep\_rem sleep\_cycle awake  
## <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 Cheet… Acin… carni Carn… lc 12.1 NA NA 11.9  
## 2 Owl m… Aotus omni Prim… <NA> 17 1.8 NA 7   
## 3 Mount… Aplo… herbi Rode… nt 14.4 2.4 NA 9.6  
## 4 Great… Blar… omni Sori… lc 14.9 2.3 0.133 9.1  
## 5 Cow Bos herbi Arti… domesticated 4 0.7 0.667 20   
## 6 Three… Brad… herbi Pilo… <NA> 14.4 2.2 0.767 9.6  
## 7 North… Call… carni Carn… vu 8.7 1.4 0.383 15.3  
## 8 Vespe… Calo… <NA> Rode… <NA> 7 NA NA 17   
## 9 Dog Canis carni Carn… domesticated 10.1 2.9 0.333 13.9  
## 10 Roe d… Capr… herbi Arti… lc 3 NA NA 21   
## # ℹ 73 more rows  
## # ℹ 2 more variables: brainwt <dbl>, bodywt <dbl>

#mutate function chnages the brainwt into 1000 times more  
 #View()

# new variable size\_of\_brain and categorizes brain weight  
size\_of\_brain <- msleep %>%   
 select(name, brainwt) %>%  
 # Select the columns 'name' and 'brainwt'  
 drop\_na(brainwt) %>%   
 # drops NA values i.e missing values  
 mutate(brain\_size = if\_else(brainwt > 0.01, "large", "small"))  
# create a new varaible name as brain\_size  
# if brainwt is more than 0.01then set its values to large else small   
print(head(size\_of\_brain))

## # A tibble: 6 × 3  
## name brainwt brain\_size  
## <chr> <dbl> <chr>   
## 1 Owl monkey 0.0155 large   
## 2 Greater short-tailed shrew 0.00029 small   
## 3 Cow 0.423 large   
## 4 Dog 0.07 large   
## 5 Roe deer 0.0982 large   
## 6 Goat 0.115 large

# prints the first few rows of the size\_of\_brain

# chnaging observations from large,small into 1,2   
size\_of\_brain %>% #   
 mutate(brain\_size = recode(brain\_size, "large" = 1, "small" = 2))

## # A tibble: 56 × 3  
## name brainwt brain\_size  
## <chr> <dbl> <dbl>  
## 1 Owl monkey 0.0155 1  
## 2 Greater short-tailed shrew 0.00029 2  
## 3 Cow 0.423 1  
## 4 Dog 0.07 1  
## 5 Roe deer 0.0982 1  
## 6 Goat 0.115 1  
## 7 Guinea pig 0.0055 2  
## 8 Chinchilla 0.0064 2  
## 9 Star-nosed mole 0.001 2  
## 10 African giant pouched rat 0.0066 2  
## # ℹ 46 more rows

# modifying brain\_size   
#recodde will first argument as which to work with   
print(head(size\_of\_brain)) # prints the head of the data frame

## # A tibble: 6 × 3  
## name brainwt brain\_size  
## <chr> <dbl> <chr>   
## 1 Owl monkey 0.0155 large   
## 2 Greater short-tailed shrew 0.00029 small   
## 3 Cow 0.423 large   
## 4 Dog 0.07 large   
## 5 Roe deer 0.0982 large   
## 6 Goat 0.115 large

# Reshapes the data from long to wide or wide to long  
library(gapminder)   
 data<- select(gapminder, country, year, lifeExp)   
 # from gap minder dataset the above code select those speicifc varaibles   
 wide\_data <- data %>%   
 # createa a new data set and pipe the data  
 pivot\_wider(names\_from = year, values\_from = lifeExp)   
 # pivot wider changes the column names as years and life expectancy as values  
 #View(wide\_data)

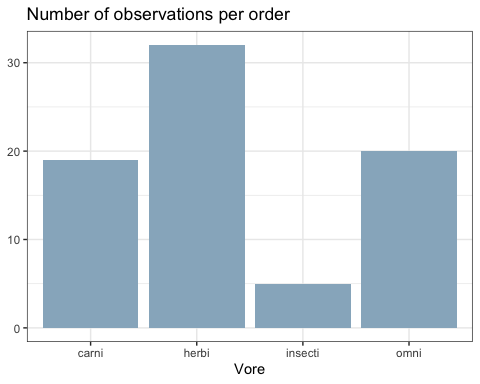
long\_data <- wide\_data %>%   
 # includes wide\_data with pipe new dataset long\_data  
pivot\_longer(2:13,   
 # slects 2 to 13 columns back from the wide data  
 names\_to = "year",   
 # pivoted columns should go into a new 'year' column  
 values\_to = "lifeExp")   
# new values to lifeExp  
#View(long\_data)

## Bar charts and Histograms using ggplot in R with Greg Martin

library(tidyverse)   
?msleep   
# displays the info about the dataframe  
head(msleep) # gets the top few rows

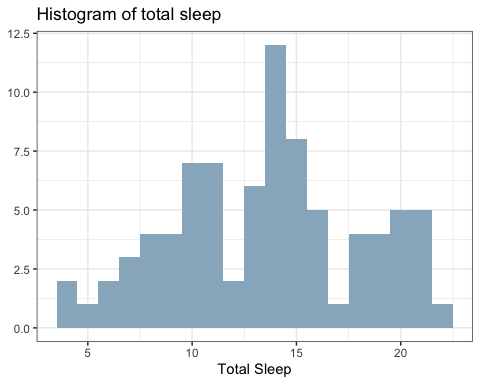
## # A tibble: 6 × 11  
## name genus vore order conservation sleep\_total sleep\_rem sleep\_cycle awake  
## <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 Cheetah Acin… carni Carn… lc 12.1 NA NA 11.9  
## 2 Owl mo… Aotus omni Prim… <NA> 17 1.8 NA 7   
## 3 Mounta… Aplo… herbi Rode… nt 14.4 2.4 NA 9.6  
## 4 Greate… Blar… omni Sori… lc 14.9 2.3 0.133 9.1  
## 5 Cow Bos herbi Arti… domesticated 4 0.7 0.667 20   
## 6 Three-… Brad… herbi Pilo… <NA> 14.4 2.2 0.767 9.6  
## # ℹ 2 more variables: brainwt <dbl>, bodywt <dbl>

msleep %>%   
 drop\_na(vore) %>%   
 # drops missing values in vore  
 ggplot(aes(x=vore))+   
 # useing ggplot and mapping vore to x axis  
 geom\_bar(fill="#97B3C6")+   
 # creates a bar plot and assigns the filler  
 theme\_bw()+   
 # sets black and white theme  
 labs(x = "Vore",   
 # labeling the x axis  
 y = NULL,  
 # labeling the y-axis as null  
 title = "Number of observations per order")



# adds the title  
# This plot gives the count in a bar graph according to carni, herbi, insects and omnivores  
#If the bars are uneven, it suggests that some categories have more species recorded than others.  
# The plot displays the count of various dietary categories ('vore') in the data set. Herbivores ('herbi') have the highest count (over 30), followed by 'carni' and 'omni', with each around 20.   
#Insectivores have the lowest count, below 5.  
#With a much taller bar, it means that most of the recorded species in the dataset belong to herbi.

# created a barplot   
msleep %>%   
 ggplot(aes(awake)) +   
 # Using ggplot with aesthetics mapping to awake  
 geom\_histogram(binwidth = 1, fill = "#97B3C6")+   
 # creates histogram with bindwidth = 1 and fills the repective color   
 theme\_bw()+   
 # sets black and white theme  
 labs(x = "Total Sleep",   
 y = NULL,   
 title = "Histogram of total sleep")



#The histogram helps identify how sleep duration varies among different species.  
# The plot shows the allocation of total sleep time in the `msleep` dataset.   
#The data shows multiple peaks, indicating variations in sleep duration across numerous species.  
#If most bars are clustered to the left (low awake hours), it suggests that many species sleep a lot.  
#Peaks indicate common sleep time among species.