Basics of programming

Jesus Sanchez

2022-10-17

#Exploring data

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.1
## v readr   2.1.1     v forcats 0.5.1

## Warning: package 'stringr' was built under R version 4.1.3

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
starwars
```

```
## # A tibble: 87 x 14
##    name    height  mass hair_color  skin_color eye_color birth_year sex   gender
##    <chr>    <int> <dbl> <chr>       <chr>      <chr>          <dbl> <chr> <chr>
##  1 Luke S~    172    77 blond       fair       blue            19   male  mascu~
##  2 C-3PO      167    75 <NA>        gold       yellow         112   none  mascu~
##  3 R2-D2       96    32 <NA>        white, bl~ red             33   none  mascu~
##  4 Darth ~    202   136 none        white      yellow          41.9 male  mascu~
##  5 Leia O~    150    49 brown       light      brown           19   fema~ femin~
##  6 Owen L~    178   120 brown, grey light      blue            52   male  mascu~
##  7 Beru W~    165    75 brown       light      blue            47   fema~ femin~
##  8 R5-D4       97    32 <NA>        white, red red             NA   none  mascu~
##  9 Biggs ~    183    84 black       light      brown           24   male  mascu~
## 10 Obi-Wa~    182    77 auburn, wh~ fair       blue-gray       57   male  mascu~
## # ... with 77 more rows, and 5 more variables: homeworld <chr>, species <chr>,
## #   films <list>, vehicles <list>, starships <list>
```

```
dim(starwars) #87 obs and 14 variables
```

```
## [1] 87 14
```

```
#str(starwars)
glimpse(starwars)
```

```
## Rows: 87
## Columns: 14
## $ name       <chr> "Luke Skywalker", "C-3PO", "R2-D2", "Darth Vader", "Leia Or~
## $ height     <int> 172, 167, 96, 202, 150, 178, 165, 97, 183, 182, 188, 180, 2~
## $ mass       <dbl> 77.0, 75.0, 32.0, 136.0, 49.0, 120.0, 75.0, 32.0, 84.0, 77.~
## $ hair_color <chr> "blond", NA, NA, "none", "brown", "brown, grey", "brown", N~
## $ skin_color <chr> "fair", "gold", "white, blue", "white", "light", "light", "~
## $ eye_color  <chr> "blue", "yellow", "red", "yellow", "brown", "blue", "blue",~
## $ birth_year <dbl> 19.0, 112.0, 33.0, 41.9, 19.0, 52.0, 47.0, NA, 24.0, 57.0, ~
## $ sex        <chr> "male", "none", "none", "male", "female", "male", "female",~
## $ gender     <chr> "masculine", "masculine", "masculine", "masculine", "femini~
## $ homeworld  <chr> "Tatooine", "Tatooine", "Naboo", "Tatooine", "Alderaan", "T~
## $ species    <chr> "Human", "Droid", "Droid", "Human", "Human", "Human", "Huma~
## $ films      <list> <"The Empire Strikes Back", "Revenge of the Sith", "Return~
## $ vehicles   <list> <"Snowspeeder", "Imperial Speeder Bike">, <>, <>, <>, "Imp~
## $ starships  <list> <"X-wing", "Imperial shuttle">, <>, <>, "TIE Advanced x1",~
```

```
head(starwars) #first 6 obs
```

```
## # A tibble: 6 x 14
##   name     height  mass hair_color  skin_color eye_color birth_year sex   gender
##   <chr>     <int> <dbl> <chr>       <chr>      <chr>          <dbl> <chr> <chr>
## 1 Luke Sk~    172    77 blond       fair       blue            19   male  mascu~
## 2 C-3PO       167    75 <NA>        gold       yellow         112   none  mascu~
## 3 R2-D2        96    32 <NA>        white, bl~ red             33   none  mascu~
## 4 Darth V~    202   136 none        white      yellow          41.9 male  mascu~
## 5 Leia Or~    150    49 brown       light      brown           19   fema~ femin~
## 6 Owen La~    178   120 brown, grey light      blue            52   male  mascu~
## # ... with 5 more variables: homeworld <chr>, species <chr>, films <list>,
## #   vehicles <list>, starships <list>
```

```
tail(starwars) #last 6 obs
```

```
## # A tibble: 6 x 14
##   name     height  mass hair_color skin_color eye_color birth_year sex    gender
##   <chr>     <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr>  <chr>
## 1 Finn         NA    NA black      dark       dark              NA male   mascu~
## 2 Rey          NA    NA brown      light      hazel             NA female femin~
## 3 Poe Dam~     NA    NA brown      light      brown             NA male   mascu~
## 4 BB8          NA    NA none       none       black             NA none   mascu~
## 5 Captain~     NA    NA unknown    unknown    unknown           NA <NA>   <NA>
## 6 Padmé A~    165    45 brown      light      brown             46 female femin~
## # ... with 5 more variables: homeworld <chr>, species <chr>, films <list>,
## #   vehicles <list>, starships <list>
```

```
attach(starwars) #this is useful to stop writing starwars$..., now I can write any variable of starwars
hair_color
```

```
##  [1] "blond"         NA              NA              "none"
##  [5] "brown"         "brown, grey"   "brown"         NA
##  [9] "black"         "auburn, white" "blond"         "auburn, grey"
## [13] "brown"         "brown"         NA              NA
## [17] "brown"         "brown"         "white"         "grey"
## [21] "black"         "none"          "none"          "black"
## [25] "none"          "none"          "auburn"        "brown"
## [29] "brown"         "none"          "brown"         "none"
## [33] "blond"         "none"          "none"          "none"
## [37] "brown"         "black"         "none"          "black"
## [41] "black"         "none"          "none"          "none"
## [45] "none"          "none"          "none"          "none"
## [49] "white"         "none"          "black"         "none"
## [53] "none"          "none"          "none"          "none"
## [57] "black"         "brown"         "brown"         "none"
## [61] "black"         "black"         "brown"         "white"
## [65] "black"         "black"         "blonde"        "none"
## [69] "none"          "none"          "white"         "none"
## [73] "none"          "none"          "none"          "none"
## [77] "none"          "brown"         "brown"         "none"
## [81] "none"          "black"         "brown"         "brown"
## [85] "none"          "unknown"       "brown"
```

```
names(starwars) #names of my variables
```

```
##  [1] "name"       "height"     "mass"       "hair_color" "skin_color"
##  [6] "eye_color"  "birth_year" "sex"        "gender"     "homeworld"
## [11] "species"    "films"      "vehicles"   "starships"
```

```
length(starwars) #for a data set length will mean the number of variables
```

```
## [1] 14
```

```
length(hair_color) #for a variable R will tell the number of obs
```

```
## [1] 87
```

```
class(hair_color)
```

```
## [1] "character"
```

```
unique(hair_color) #name of unique obs
```

```
##  [1] "blond"         NA              "none"          "brown"
##  [5] "brown, grey"   "black"         "auburn, white" "auburn, grey"
##  [9] "white"         "grey"          "auburn"        "blonde"
## [13] "unknown"
```

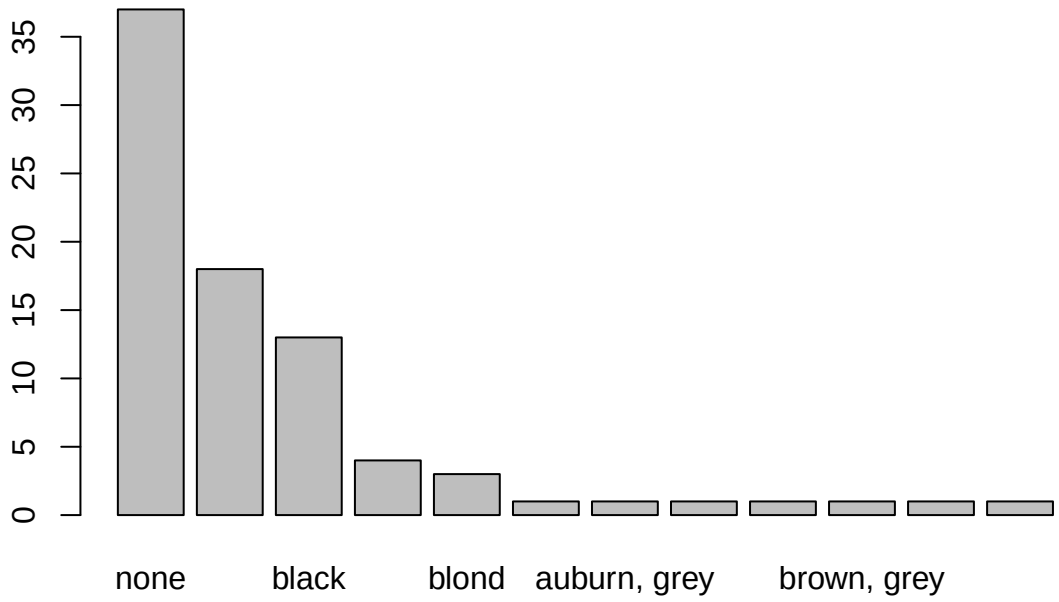
```r
#na: data is missing
#none: hair without a color or there´s no hair
#unknow: we don't know, maybe the character uses a hat, so we don't know the color

table(hair_color)
```

```
## hair_color
##        auburn  auburn, grey auburn, white         black         blond
##             1             1             1            13             3
##        blonde         brown   brown, grey          grey          none
##             1            18             1             1            37
##       unknown         white
##             1             4
```

```r
sort(table(hair_color), decreasing=T)
```

```
## hair_color
##          none         brown         black         white         blond
##            37            18            13             4             3
##        auburn  auburn, grey auburn, white        blonde   brown, grey
##             1             1             1             1             1
##          grey       unknown
##             1             1
```

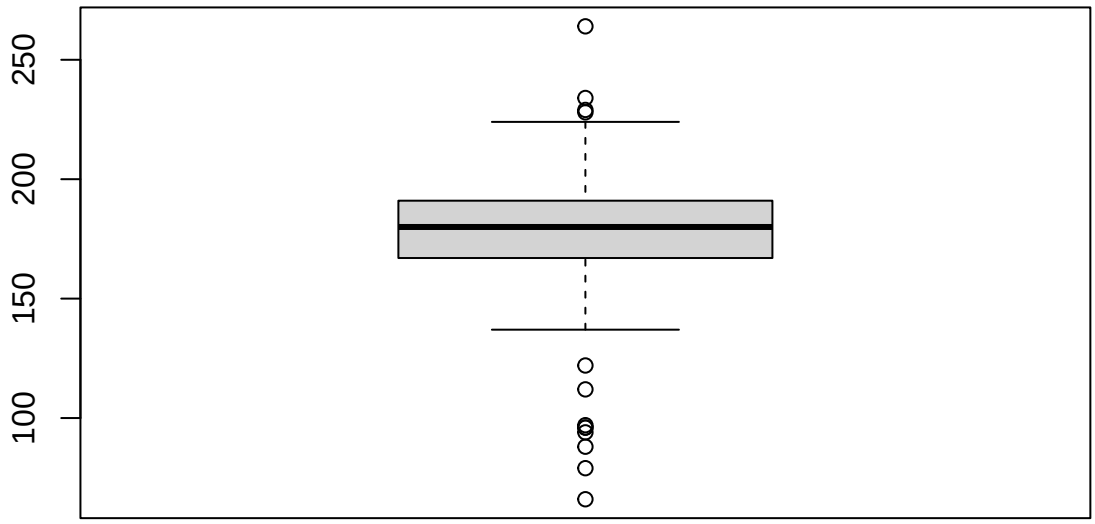
```r
View(sort(table(hair_color), decreasing=T))
barplot(sort(table(hair_color), decreasing=T))
```



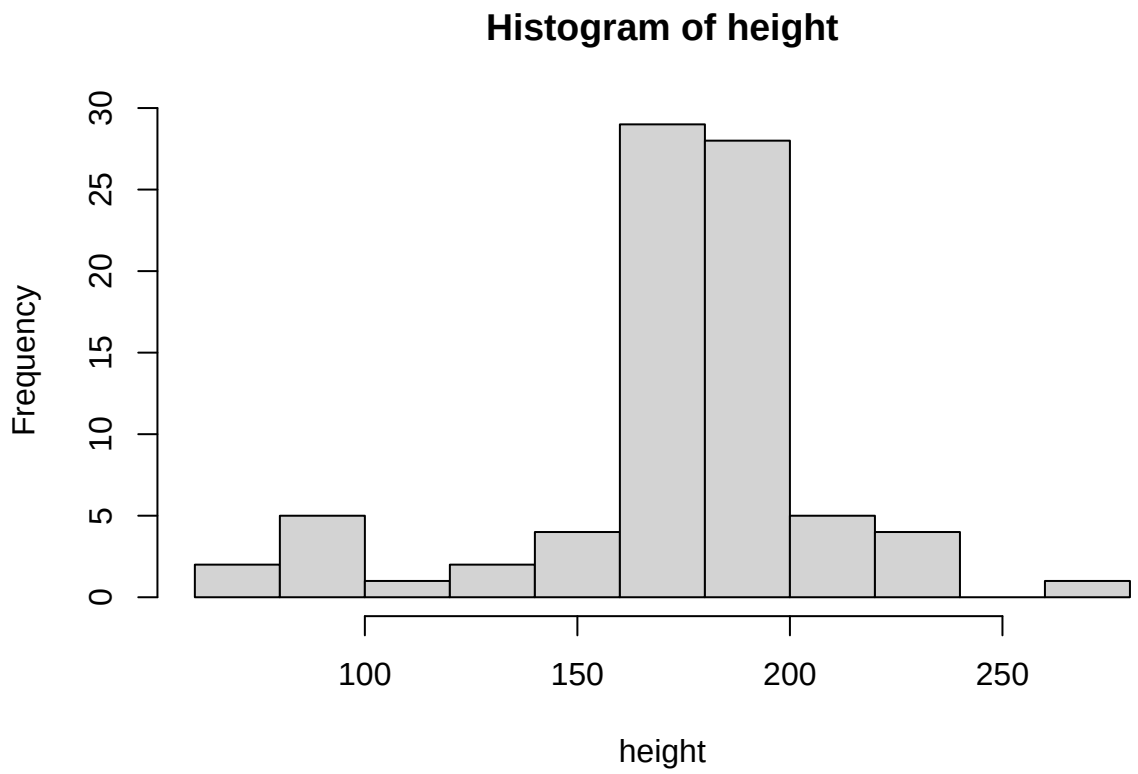

```
#pipes operators

starwars %>%
  select(hair_color) %>%
  count(hair_color) %>%
  arrange(desc(n)) %>%
  View()

View(starwars[is.na(hair_color),])  #selecting row where is.na is TRUE

summary(height)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    66.0   167.0   180.0   174.4   191.0   264.0       6
```

```
boxplot(height) #boxplot
```




```
hist(height) #histeogram
```



#Cleaning data

```
library(tidyverse)
data()
View(starwars)
glimpse(starwars)
```

```
## Rows: 87
## Columns: 14
## $ name       <chr> "Luke Skywalker", "C-3PO", "R2-D2", "Darth Vader", "Leia Or~
## $ height     <int> 172, 167, 96, 202, 150, 178, 165, 97, 183, 182, 188, 180, 2~
## $ mass       <dbl> 77.0, 75.0, 32.0, 136.0, 49.0, 120.0, 75.0, 32.0, 84.0, 77.~
## $ hair_color <chr> "blond", NA, NA, "none", "brown", "brown, grey", "brown", N~
## $ skin_color <chr> "fair", "gold", "white, blue", "white", "light", "light", "~
## $ eye_color  <chr> "blue", "yellow", "red", "yellow", "brown", "blue", "blue",~
## $ birth_year <dbl> 19.0, 112.0, 33.0, 41.9, 19.0, 52.0, 47.0, NA, 24.0, 57.0, ~
## $ sex        <chr> "male", "none", "none", "male", "female", "male", "female",~
## $ gender     <chr> "masculine", "masculine", "masculine", "masculine", "femini~
## $ homeworld  <chr> "Tatooine", "Tatooine", "Naboo", "Tatooine", "Alderaan", "T~
## $ species    <chr> "Human", "Droid", "Droid", "Human", "Human", "Human", "Huma~
## $ films      <list> <"The Empire Strikes Back", "Revenge of the Sith", "Return~
## $ vehicles   <list> <"Snowspeeder", "Imperial Speeder Bike">, <>, <>, <>, "Imp~
## $ starships  <list> <"X-wing", "Imperial shuttle">, <>, <>, "TIE Advanced x1",~
```

```r
unique(starwars$gender) #displays data type in the obs of a specific column
```

```
## [1] "masculine" "feminine"  NA
```

```r
starwars$gender <- as.factor(starwars$gender)
class(starwars$gender)#now gender is a factor
```

```
## [1] "factor"
```

```r
levels(starwars$gender)
```

```
## [1] "feminine"  "masculine"
```

```r
starwars$gender <- factor((starwars$gender), levels = c("feminine", "masculine"))
#changing levels

starwars %>% select(name, height, ends_with("color")) %>%
  names()
```

```
## [1] "name"       "height"     "hair_color" "skin_color" "eye_color"
```

```r
unique(starwars$hair_color)
```

```
##  [1] "blond"         NA              "none"          "brown"
##  [5] "brown, grey"   "black"         "auburn, white" "auburn, grey"
##  [9] "white"         "grey"          "auburn"        "blonde"
## [13] "unknown"
```

```r
starwars %>%
  select(name, height, ends_with("color")) %>%
  filter(hair_color %in% c("blond", "brown") & height < 180)
```

```
## # A tibble: 9 x 5
##   name                  height hair_color skin_color eye_color
##   <chr>                  <int> <chr>      <chr>      <chr>
## 1 Luke Skywalker           172 blond      fair       blue
## 2 Leia Organa              150 brown      light      brown
## 3 Beru Whitesun lars       165 brown      light      blue
## 4 Wedge Antilles           170 brown      fair       hazel
## 5 Wicket Systri Warrick     88 brown      brown      brown
## 6 Finis Valorum            170 blond      fair       blue
## 7 Cordé                    157 brown      light      brown
## 8 Dormé                    165 brown      light      brown
## 9 Padmé Amidala            165 brown      light      brown
```

```r
#%in% works for group more than 1 variable

#missing data

mean(starwars$height) #we have a NA because there's missin values Na
```

```
## [1] NA
```

```
mean(starwars$height, na.rm = T)
```

```
## [1] 174.358
```

```
starwars %>%
  select(name, gender, hair_color, height) %>%
  na.omit()
```

```
## # A tibble: 73 x 4
##    name               gender    hair_color    height
##    <chr>              <fct>     <chr>          <int>
##  1 Luke Skywalker     masculine blond            172
##  2 Darth Vader        masculine none             202
##  3 Leia Organa        feminine  brown            150
##  4 Owen Lars          masculine brown, grey      178
##  5 Beru Whitesun lars feminine  brown            165
##  6 Biggs Darklighter  masculine black            183
##  7 Obi-Wan Kenobi     masculine auburn, white    182
##  8 Anakin Skywalker   masculine blond            188
##  9 Wilhuff Tarkin     masculine auburn, grey     180
## 10 Chewbacca          masculine brown            228
## # ... with 63 more rows
```

```
starwars %>%
  select(name, gender, hair_color, height) %>%
  filter(!complete.cases(.)) #what obs we deleted
```

```
## # A tibble: 14 x 4
##    name                  gender    hair_color height
##    <chr>                 <fct>     <chr>       <int>
##  1 C-3PO                 masculine <NA>          167
##  2 R2-D2                 masculine <NA>           96
##  3 R5-D4                 masculine <NA>           97
##  4 Greedo                masculine <NA>          173
##  5 Jabba Desilijic Tiure masculine <NA>          175
##  6 Arvel Crynyd          masculine brown          NA
##  7 Ric Olié              <NA>      brown         183
##  8 Quarsh Panaka         <NA>      black         183
##  9 Sly Moore             <NA>      none          178
## 10 Finn                  masculine black          NA
## 11 Rey                   feminine  brown          NA
## 12 Poe Dameron           masculine brown          NA
## 13 BB8                   masculine none           NA
## 14 Captain Phasma        <NA>      unknown        NA
```

```
starwars %>%
  select(name, gender, hair_color, height) %>%
  filter(!complete.cases(.)) %>%
  drop_na(height)
```

```
## # A tibble: 8 x 4
##   name                  gender    hair_color height
##   <chr>                 <fct>     <chr>       <int>
## 1 C-3PO                 masculine <NA>          167
## 2 R2-D2                 masculine <NA>           96
## 3 R5-D4                 masculine <NA>           97
## 4 Greedo                masculine <NA>          173
## 5 Jabba Desilijic Tiure masculine <NA>          175
## 6 Ric Olié              <NA>      brown         183
## 7 Quarsh Panaka         <NA>      black         183
## 8 Sly Moore             <NA>      none          178
```

```
starwars %>%
  select(name, gender, hair_color, height) %>%
  filter(!complete.cases(.)) %>%
  mutate(hair_color = replace_na(hair_color, "none"))
```

```
## # A tibble: 14 x 4
##    name                  gender    hair_color height
##    <chr>                 <fct>     <chr>       <int>
##  1 C-3PO                 masculine none          167
##  2 R2-D2                 masculine none           96
##  3 R5-D4                 masculine none           97
##  4 Greedo                masculine none          173
##  5 Jabba Desilijic Tiure masculine none          175
##  6 Arvel Crynyd          masculine brown          NA
##  7 Ric Olié              <NA>      brown         183
##  8 Quarsh Panaka         <NA>      black         183
##  9 Sly Moore             <NA>      none          178
## 10 Finn                  masculine black          NA
## 11 Rey                   feminine  brown          NA
## 12 Poe Dameron           masculine brown          NA
## 13 BB8                   masculine none           NA
## 14 Captain Phasma        <NA>      unknown        NA
```

```
#replacing all NA values from hair_color

#Duplicates-------------------------------

Names <- c("Peter", "John", "Andrew", "Peter")
Age <- c(22,33,44,22)

friends <- data.frame(Names, Age)
duplicated(friends) #reporting duplicates
```

```
## [1] FALSE FALSE FALSE  TRUE
```

```
friends[!duplicated(friends), ] #the archaic method
```

```
##    Names Age
## 1  Peter  22
## 2   John  33
## 3 Andrew  44
```

```r
friends %>% distinct() #using tydiverse
```

```
##    Names Age
## 1  Peter  22
## 2   John  33
## 3 Andrew  44
```

```r
#recording variables----------------------------------

starwars %>% select(name, gender)
```

```
## # A tibble: 87 x 2
##    name               gender
##    <chr>              <fct>
##  1 Luke Skywalker     masculine
##  2 C-3PO              masculine
##  3 R2-D2              masculine
##  4 Darth Vader        masculine
##  5 Leia Organa        feminine
##  6 Owen Lars          masculine
##  7 Beru Whitesun lars feminine
##  8 R5-D4              masculine
##  9 Biggs Darklighter  masculine
## 10 Obi-Wan Kenobi     masculine
## # ... with 77 more rows
```

```r
class(starwars$gender)
```

```
## [1] "factor"
```

```r
starwars$gender <- as.factor(starwars$gender)
class(starwars$gender) #now we can recode the variable
```

```
## [1] "factor"
```

```r
levels(starwars$gender)
```

```
## [1] "feminine"  "masculine"
```

```r
starwars %>%
  select(name, gender) %>%
  mutate(gender_coded = recode(gender,
                         "masculine"= 1,
                         "feminine" = 2))
```

```
## # A tibble: 87 x 3
##    name               gender    gender_coded
##    <chr>              <fct>            <dbl>
##  1 Luke Skywalker     masculine            1
```

```
##  2 C-3PO              masculine           1
##  3 R2-D2              masculine           1
##  4 Darth Vader        masculine           1
##  5 Leia Organa        feminine            2
##  6 Owen Lars          masculine           1
##  7 Beru Whitesun lars feminine            2
##  8 R5-D4              masculine           1
##  9 Biggs Darklighter  masculine           1
## 10 Obi-Wan Kenobi     masculine           1
## # ... with 77 more rows
```

#Manipulating data

```
library(tidyverse)

?msleep
```

```
## starting httpd help server ... done
```

```
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~
```

```
#rename a variable

msleep %>%
  rename("conserv"= "conservation") %>%
  glimpse()
```

```
## Rows: 83
## Columns: 11
## $ name        <chr> "Cheetah", "Owl monkey", "Mountain beaver", "Greater short~
## $ genus       <chr> "Acinonyx", "Aotus", "Aplodontia", "Blarina", "Bos", "Brad~
## $ vore        <chr> "carni", "omni", "herbi", "omni", "herbi", "herbi", "carni~
## $ order       <chr> "Carnivora", "Primates", "Rodentia", "Soricomorpha", "Arti~
## $ conserv     <chr> "lc", NA, "nt", "lc", "domesticated", NA, "vu", NA, "domes~
## $ 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, 0~
## $ sleep_cycle <dbl> NA, NA, NA, 0.1333333, 0.6666667, 0.7666667, 0.3833333, NA~
## $ awake       <dbl> 11.9, 7.0, 9.6, 9.1, 20.0, 9.6, 15.3, 17.0, 13.9, 21.0, 18~
## $ 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, 0.045~
```

```r
#reorder a variable

msleep %>%
  select(vore, name, everything())
```

```
## # A tibble: 83 x 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
## # ... with 73 more rows, and 2 more variables: brainwt <dbl>, bodywt <dbl>
```

```r
#Change a variable name

class(msleep$vore)
```

```
## [1] "character"
```

```r
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~
```

```r
msleep$vore <- as.factor(msleep$vore)
glimpse(msleep$vore)
```

```
##  Factor w/ 4 levels "carni","herbi",..: 1 4 2 4 2 2 1 NA 1 2 ...
```

```r
msleep %>%
  mutate(vore = as.character(vore)) %>%
  glimpse()
```

```
## 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~
```

```
#select variables to work with

names(msleep)
```

```
##  [1] "name"         "genus"        "vore"         "order"        "conservation"
##  [6] "sleep_total"  "sleep_rem"    "sleep_cycle"  "awake"        "brainwt"
## [11] "bodywt"
```

```
msleep %>%
  select(2:4,
         awake,
         starts_with("sleep"),
         contains("wt")) %>%
  names()
```

```
## [1] "genus"       "vore"        "order"       "awake"       "sleep_total"
## [6] "sleep_rem"   "sleep_cycle" "brainwt"     "bodywt"
```

```
#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"
```

```
msleep %>%
  filter((order=="Carnivora" |
          order=="Primates") &
           sleep_total > 8) %>%
  select(name, order, sleep_total) %>%
  arrange(-sleep_total) %>%
  View

msleep %>%
  filter(order %in% c("Carnivora","Primates") &
```

```
          sleep_total > 8) %>%
  select(name, order, sleep_total) %>%
  arrange(order) %>%
  View

#change observations (mutate)

msleep %>%
  mutate(brainwt_grams = brainwt * 1000) %>%
  View

#conditional changes(if_else)
#logical vector based on a conditional

msleep$brainwt > 0.01
```

```
##  [1]    NA  TRUE    NA FALSE  TRUE    NA    NA    NA  TRUE  TRUE  TRUE FALSE
## [13]    NA FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE  TRUE FALSE  TRUE  TRUE
## [25] FALSE  TRUE    NA  TRUE FALSE    NA    NA  TRUE  TRUE  TRUE    NA  TRUE
## [37]    NA  TRUE    NA FALSE    NA FALSE FALSE    NA  TRUE    NA    NA  TRUE
## [49]  TRUE  TRUE    NA  TRUE    NA  TRUE FALSE    NA    NA  TRUE    NA    NA
## [61]    NA  TRUE  TRUE FALSE    NA  TRUE FALSE FALSE FALSE FALSE FALSE    NA
## [73] FALSE  TRUE  TRUE    NA  TRUE FALSE FALSE    NA  TRUE  TRUE  TRUE
```

```
size_of_brain <- msleep %>%
  select(name, brainwt) %>%
  drop_na(brainwt) %>%
  mutate(brain_size = if_else(brainwt > 0.01,
                              "large",
                              "small"))
  View #If it is true then large, if not then small
```

```
## function (x, title)
## {
##     check <- Sys.getenv("_R_CHECK_SCREEN_DEVICE_", "")
##     msg <- "View() should not be used in examples etc"
##     if (identical(check, "stop"))
##         stop(msg, domain = NA)
##     else if (identical(check, "warn"))
##         warning(msg, immediate. = TRUE, noBreaks. = TRUE, domain = NA)
##     if (missing(title))
##         title <- paste("Data:", deparse(substitute(x))[1])
##     x0 <- as.data.frame(x)
##     x <- as.list(format.data.frame(x0))
##     rn <- row.names(x0)
##     if (any(rn != seq_along(rn)))
##         x <- c(list(row.names = rn), x)
##     if (!is.list(x) || !length(x) || !all(sapply(x, is.atomic)) ||
##         !max(lengths(x)))
##         stop("invalid 'x' argument")
##     if (grepl("darwin", R.version$os))
##         check_for_XQuartz()
```

```
##     invisible(.External2(C_dataviewer, x, title))
## }
## <bytecode: 0x0000000020d95738>
## <environment: namespace:utils>
```

```
#Recode data and rename a variable
##Change obs of "large" and "small" into

size_of_brain %>%
  mutate(brain_size = recode(brain_size,
                             "large"= 1,
                             "small"=2))
```

```
## # A tibble: 56 x 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
## # ... with 46 more rows
```

```
#reshape the data from wide to long or long to wide


library(gapminder)
```

```
## Warning: package 'gapminder' was built under R version 4.1.3
```

```
View(gapminder)

data <- select(gapminder, country, year, lifeExp)
data
```

```
## # A tibble: 1,704 x 3
##    country      year lifeExp
##    <fct>       <int>   <dbl>
##  1 Afghanistan  1952    28.8
##  2 Afghanistan  1957    30.3
##  3 Afghanistan  1962    32.0
##  4 Afghanistan  1967    34.0
##  5 Afghanistan  1972    36.1
##  6 Afghanistan  1977    38.4
##  7 Afghanistan  1982    39.9
##  8 Afghanistan  1987    40.8
##  9 Afghanistan  1992    41.7
## 10 Afghanistan  1997    41.8
## # ... with 1,694 more rows
```

#Visualise

#Analyse