

## R intensive: Introduction to dplyr (Session 2)

```
knitr::opts_chunk$set(echo = TRUE)
# global options
```

```
library('tidyverse')
library('dplyr')
```

We can look at all of the data sets available to us:

```
data()
```

Here is a view of the data without printing out the whole thing:

```
glimpse(starwars)
```

```
## Rows: 87
## Columns: 14
## $ name      <chr> "Luke Skywalker", "C-3P0", "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",~
```

## Column operations

Manipulate the columns (variables) of your data frame:

### 1. Using the select function to select specific columns in a data frame (tibble)

```
starwars.cut <- select(starwars, c(1:6, 8:10),) #dplyr
```

```
starwars.cut <- starwars[, c(1:6, 8:10)] #base R
```

```
glimpse(starwars.cut)
```

```
## Rows: 87
## Columns: 9
## $ name      <chr> "Luke Skywalker", "C-3P0", "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",~
## $ sex        <chr> "male", "none", "none", "male", "female", "male", "female",~
## $ gender     <chr> "masculine", "masculine", "masculine", "masculine", "femini~
## $ homeworld  <chr> "Tatooine", "Tatooine", "Naboo", "Tatooine", "Alderaan", "T~
```

```
head(starwars.cut)
```

```
## # A tibble: 6 x 9
##   name      height  mass hair_color skin_color eye_color sex  gender homeworld
##   <chr>      <int> <dbl> <chr>      <chr>      <chr> <chr> <chr> <chr>
## 1 Luke Skyw~   172    77 blond      fair        blue   male  mascul~ Tatooine
## 2 C-3P0       167    75 <NA>       gold        yellow none  mascul~ Tatooine
## 3 R2-D2       96     32 <NA>       white, bl~ red     none  mascul~ Naboo
## 4 Darth Vad~  202   136 none       white       yellow male  mascul~ Tatooine
## 5 Leia Orga~  150    49 brown      light       brown  fema~ femin~ Alderaan
## 6 Owen Lars   178   120 brown, gr~ light       blue   male  mascul~ Tatooine
```

**FYI** - There are some differences in style when coding. For example, these are the same thing:

```
starwars.cut <- starwars[,c(1:6,8:10)]
starwars.cut = starwars[,c(1:6,8:10)]
```

*However, it is considered better form to use the assignment operator <-*

**Naming convention:** We can use either an underscore (`_`) or period (`.`) in naming our variables and data sets. Both are fine, but stick to one approach.

```
starwars.cut <- starwars[,c(1:6,8:10)]
```

is not the same data frame as

```
starwars_cut <- starwars[,c(1:6,8:10)]
```

## 2. Using the mutate function - to manipulate a variable

#Suppose we want to compute height in feet instead of meters

```
mutate(starwars.cut, height.feet = height/30.48)
```

```
## # A tibble: 87 x 10
##   name      height  mass hair_color skin_color eye_color sex  gender homeworld
##   <chr>      <int> <dbl> <chr>      <chr>      <chr> <chr> <chr> <chr>
## 1 Luke Skyw~   172    77 blond      fair        blue   male  mascul~ Tatooine
## 2 C-3P0       167    75 <NA>       gold        yellow none  mascul~ Tatooine
## 3 R2-D2       96     32 <NA>       white, bl~ red     none  mascul~ Naboo
```

```
## 4 Darth Va~ 202 136 none white yellow male mascu~ Tatooine
## 5 Leia Org~ 150 49 brown light brown fema~ femin~ Alderaan
## 6 Owen Lars 178 120 brown, gr~ light blue male mascu~ Tatooine
## 7 Beru Whi~ 165 75 brown light blue fema~ femin~ Tatooine
## 8 R5-D4 97 32 <NA> white, red red none mascu~ Tatooine
## 9 Biggs Da~ 183 84 black light brown male mascu~ Tatooine
## 10 Obi-Wan ~ 182 77 auburn, w~ fair blue-gray male mascu~ Stewjon
## # i 77 more rows
## # i 1 more variable: height.feet <dbl>
```

### 3. Combining verbs with pipes (%>%)

```
mutate(starwars.cut, height.feet = height/30.48) # Without pipes
```

```
## # A tibble: 87 x 10
##   name      height  mass hair_color skin_color eye_color sex  gender homeworld
##   <chr>    <int> <dbl> <chr>    <chr>    <chr> <chr> <chr> <chr>
## 1 Luke Sky~ 172 77 blond fair blue male mascu~ Tatooine
## 2 C-3PO 167 75 <NA> gold yellow none mascu~ Tatooine
## 3 R2-D2 96 32 <NA> white, bl~ red none mascu~ Naboo
## 4 Darth Va~ 202 136 none white yellow male mascu~ Tatooine
## 5 Leia Org~ 150 49 brown light brown fema~ femin~ Alderaan
## 6 Owen Lars 178 120 brown, gr~ light blue male mascu~ Tatooine
## 7 Beru Whi~ 165 75 brown light blue fema~ femin~ Tatooine
## 8 R5-D4 97 32 <NA> white, red red none mascu~ Tatooine
## 9 Biggs Da~ 183 84 black light brown male mascu~ Tatooine
## 10 Obi-Wan ~ 182 77 auburn, w~ fair blue-gray male mascu~ Stewjon
## # i 77 more rows
## # i 1 more variable: height.feet <dbl>
```

```
starwars.cut %>% mutate(height.feet = height/30.48) # With pipes
```

```
## # A tibble: 87 x 10
##   name      height  mass hair_color skin_color eye_color sex  gender homeworld
##   <chr>    <int> <dbl> <chr>    <chr>    <chr> <chr> <chr> <chr>
## 1 Luke Sky~ 172 77 blond fair blue male mascu~ Tatooine
## 2 C-3PO 167 75 <NA> gold yellow none mascu~ Tatooine
## 3 R2-D2 96 32 <NA> white, bl~ red none mascu~ Naboo
## 4 Darth Va~ 202 136 none white yellow male mascu~ Tatooine
## 5 Leia Org~ 150 49 brown light brown fema~ femin~ Alderaan
## 6 Owen Lars 178 120 brown, gr~ light blue male mascu~ Tatooine
## 7 Beru Whi~ 165 75 brown light blue fema~ femin~ Tatooine
## 8 R5-D4 97 32 <NA> white, red red none mascu~ Tatooine
## 9 Biggs Da~ 183 84 black light brown male mascu~ Tatooine
## 10 Obi-Wan ~ 182 77 auburn, w~ fair blue-gray male mascu~ Stewjon
## # i 77 more rows
## # i 1 more variable: height.feet <dbl>
```

#### 4. Renaming columns with the rename function

```
rename(starwars.cut,eye.color=eye_color)
```

```
## # A tibble: 87 x 9
##   name      height  mass hair_color skin_color eye_color sex  gender homeworld
##   <chr>    <int> <dbl> <chr>    <chr>    <chr>  <chr> <chr> <chr>
## 1 Luke Sky~  172    77 blond    fair      blue   male  mascu~ Tatooine
## 2 C-3PO     167    75 <NA>     gold      yellow none  mascu~ Tatooine
## 3 R2-D2     96     32 <NA>     white, bl~ red     none  mascu~ Naboo
## 4 Darth Va~ 202   136 none     white     yellow male  mascu~ Tatooine
## 5 Leia Org~ 150    49 brown    light     brown  fema~ femin~ Alderaan
## 6 Owen Lars 178   120 brown, gr~ light     blue   male  mascu~ Tatooine
## 7 Beru Whi~ 165    75 brown    light     blue   fema~ femin~ Tatooine
## 8 R5-D4     97     32 <NA>     white, red red     none  mascu~ Tatooine
## 9 Biggs Da~ 183    84 black    light     brown  male  mascu~ Tatooine
## 10 Obi-Wan ~ 182    77 auburn, w~ fair      blue-gray male  mascu~ Stewjon
## # i 77 more rows
```

```
starwars.upper <- rename_with(starwars.cut,toupper)
starwars.upper
```

```
## # A tibble: 87 x 9
##   NAME      HEIGHT  MASS HAIR_COLOR SKIN_COLOR EYE_COLOR SEX  GENDER HOMEWORLD
##   <chr>    <int> <dbl> <chr>    <chr>    <chr>  <chr> <chr> <chr>
## 1 Luke Sky~  172    77 blond    fair      blue   male  mascu~ Tatooine
## 2 C-3PO     167    75 <NA>     gold      yellow none  mascu~ Tatooine
## 3 R2-D2     96     32 <NA>     white, bl~ red     none  mascu~ Naboo
## 4 Darth Va~ 202   136 none     white     yellow male  mascu~ Tatooine
## 5 Leia Org~ 150    49 brown    light     brown  fema~ femin~ Alderaan
## 6 Owen Lars 178   120 brown, gr~ light     blue   male  mascu~ Tatooine
## 7 Beru Whi~ 165    75 brown    light     blue   fema~ femin~ Tatooine
## 8 R5-D4     97     32 <NA>     white, red red     none  mascu~ Tatooine
## 9 Biggs Da~ 183    84 black    light     brown  male  mascu~ Tatooine
## 10 Obi-Wan ~ 182    77 auburn, w~ fair      blue-gray male  mascu~ Stewjon
## # i 77 more rows
```

#### 5. Relocate columns with relocate function

```
relocate(starwars.upper,MASS,.after=last_col())
```

```
## # A tibble: 87 x 9
##   NAME      HEIGHT HAIR_COLOR SKIN_COLOR EYE_COLOR SEX  GENDER HOMEWORLD  MASS
##   <chr>    <int> <chr>    <chr>    <chr>    <chr> <chr> <chr>    <dbl>
## 1 Luke Sky~  172 blond    fair      blue   male  mascu~ Tatooine    77
## 2 C-3PO     167 <NA>     gold      yellow  none  mascu~ Tatooine    75
## 3 R2-D2     96 <NA>     white, bl~ red     none  mascu~ Naboo      32
## 4 Darth Va~ 202 none     white     yellow  male  mascu~ Tatooine   136
## 5 Leia Org~ 150 brown    light     brown  fema~ femin~ Alderaan    49
```

```
## 6 Owen Lars      178 brown, gr~ light      blue      male mascu~ Tatooine      120
## 7 Beru Whi~      165 brown      light      blue      fema~ femin~ Tatooine      75
## 8 R5-D4           97 <NA>         white, red red      none mascu~ Tatooine      32
## 9 Biggs Da~      183 black      light      brown     male mascu~ Tatooine      84
## 10 Obi-Wan ~     182 auburn, w~ fair      blue-gray male mascu~ Stewjon      77
## # i 77 more rows
```

## 6. Extract a column with the *pull* function

```
pull(starwars.cut,height)
```

```
## [1] 172 167 96 202 150 178 165 97 183 182 188 180 228 180 173 175 170 180 66
## [20] 170 183 200 190 177 175 180 150 NA 88 160 193 191 170 196 224 206 183 137
## [39] 112 183 163 175 180 178 94 122 163 188 198 196 171 184 188 264 188 196 185
## [58] 157 183 183 170 166 165 193 191 183 168 198 229 213 167 79 96 193 191 178
## [77] 216 234 188 178 206 NA NA NA NA NA 165
```

```
pull(starwars.cut,1)
```

```
## [1] "Luke Skywalker"      "C-3PO"          "R2-D2"
## [4] "Darth Vader"         "Leia Organa"    "Owen Lars"
## [7] "Beru Whitesun lars"  "R5-D4"          "Biggs Darklighter"
## [10] "Obi-Wan Kenobi"     "Anakin Skywalker" "Wilhuff Tarkin"
## [13] "Chewbacca"          "Han Solo"        "Greedo"
## [16] "Jabba Desilijic Tiure" "Wedge Antilles" "Jek Tono Porkins"
## [19] "Yoda"                "Palpatine"      "Boba Fett"
## [22] "IG-88"              "Bossk"          "Lando Calrissian"
## [25] "Lobot"              "Ackbar"         "Mon Mothma"
## [28] "Arvel Crynyd"       "Wicket Systri Warrick" "Nien Nunb"
## [31] "Qui-Gon Jinn"       "Nute Gunray"    "Finis Valorum"
## [34] "Jar Jar Binks"     "Roos Tarpals"   "Rugor Nass"
## [37] "Ric Olié"           "Watto"          "Sebulba"
## [40] "Quarsh Panaka"     "Shmi Skywalker" "Darth Maul"
## [43] "Bib Fortuna"       "Ayla Secura"    "Dud Bolt"
## [46] "Gasgano"           "Ben Quadinaros" "Mace Windu"
## [49] "Ki-Adi-Mundi"     "Kit Fisto"      "Eeth Koth"
## [52] "Adi Gallia"        "Saesee Tiin"   "Yarael Poof"
## [55] "Plo Koon"          "Mas Amedda"     "Gregar Typho"
## [58] "Cordé"             "Cliegg Lars"   "Poggle the Lesser"
## [61] "Luminara Unduli"   "Barriss Offee"  "Dormé"
## [64] "Dooku"             "Bail Prestor Organa" "Jango Fett"
## [67] "Zam Wesell"        "Dexter Jettster" "Lama Su"
## [70] "Taun We"           "Jocasta Nu"     "Ratts Tyerell"
## [73] "R4-P17"            "Wat Tambor"     "San Hill"
## [76] "Shaak Ti"          "Grievous"       "Tarfful"
## [79] "Raymus Antilles"   "Sly Moore"      "Tion Medon"
## [82] "Finn"              "Rey"            "Poe Dameron"
## [85] "BB8"               "Captain Phasma" "Padmé Amidala"
```

```
select(starwars,c(1:2),) #Think of the select function as a more general version of pull
```

```
## # A tibble: 87 x 2
##   name          height
##   <chr>         <int>
## 1 Luke Skywalker    172
## 2 C-3PO             167
## 3 R2-D2             96
## 4 Darth Vader      202
## 5 Leia Organa      150
## 6 Owen Lars        178
## 7 Beru Whitesun lars 165
## 8 R5-D4             97
## 9 Biggs Darklighter 183
## 10 Obi-Wan Kenobi   182
## # i 77 more rows
```

## Row operations

Manipulate the rows (observations) in your data frame

### 1. Order the rows of a data frame by the values of selected columns with the arrange function

```
starwars.cut %>% arrange(desc(height))
```

```
## # A tibble: 87 x 9
##   name          height mass hair_color skin_color eye_color sex  gender homeworld
##   <chr>         <int> <dbl> <chr>    <chr>    <chr>  <chr> <chr> <chr>
## 1 Yarael P~     264    NA none    white    yellow male  mascu~ Quermia
## 2 Tarfful      234   136 brown   brown    blue    male  mascu~ Kashyyyk
## 3 Lama Su      229    88 none    grey     black   male  mascu~ Kamino
## 4 Chewbacca    228   112 brown   unknown  blue    male  mascu~ Kashyyyk
## 5 Roos Tar~    224    82 none    grey     orange  male  mascu~ Naboo
## 6 Grievous     216   159 none    brown, wh~ green, y~ male  mascu~ Kalee
## 7 Taun We      213    NA none    grey     black   fema~ femin~ Kamino
## 8 Rugor Na~    206    NA none    green    orange  male  mascu~ Naboo
## 9 Tion Med~    206    80 none    grey     black   male  mascu~ Utapau
## 10 Darth Va~    202   136 none    white    yellow  male  mascu~ Tatooine
## # i 77 more rows
```

```
arrange(starwars.cut, desc(height))
```

```
## # A tibble: 87 x 9
##   name          height mass hair_color skin_color eye_color sex  gender homeworld
##   <chr>         <int> <dbl> <chr>    <chr>    <chr>  <chr> <chr> <chr>
## 1 Yarael P~     264    NA none    white    yellow male  mascu~ Quermia
## 2 Tarfful      234   136 brown   brown    blue    male  mascu~ Kashyyyk
## 3 Lama Su      229    88 none    grey     black   male  mascu~ Kamino
```

```
## 4 Chewbacca      228   112 brown      unknown   blue      male  mascu~ Kashyyyk
## 5 Roos Tar~     224    82 none       grey      orange   male  mascu~ Naboo
## 6 Grievous      216   159 none       brown, wh~ green, y~ male  mascu~ Kalee
## 7 Taun We       213    NA none       grey      black    fema~ femin~ Kamino
## 8 Rugor Na~     206    NA none       green     orange   male  mascu~ Naboo
## 9 Tion Med~     206    80 none       grey      black    male  mascu~ Utapau
## 10 Darth Va~   202   136 none       white     yellow   male  mascu~ Tatooine
## # i 77 more rows
```

*# Order by more than one variable. Here we are ordering by gender and then by height from tallest to s*

```
starwars.cut %>% arrange(gender, desc(height))
```

```
## # A tibble: 87 x 9
##   name      height  mass hair_color skin_color eye_color sex  gender homeworld
##   <chr>    <int> <dbl> <chr>    <chr>    <chr>  <chr> <chr> <chr>
## 1 Taun We      213  NA   none     grey     black  fema~ femin~ Kamino
## 2 Adi Gall~    184  50   none     dark     blue   fema~ femin~ Coruscant
## 3 Ayla Sec~    178  55   none     blue     hazel  fema~ femin~ Ryloth
## 4 Shaak Ti     178  57   none     red, blue~ black  fema~ femin~ Shili
## 5 Luminara~   170  56.2 black   yellow   blue   fema~ femin~ Mirial
## 6 Zam Wese~   168  55   blonde  fair, gre~ yellow  fema~ femin~ Zolan
## 7 Jocasta ~   167  NA   white    fair     blue   fema~ femin~ Coruscant
## 8 Barriss ~   166  50   black    yellow   blue   fema~ femin~ Mirial
## 9 Beru Whi~   165  75   brown    light    blue   fema~ femin~ Tatooine
## 10 Dormé      165  NA   brown    light    brown  fema~ femin~ Naboo
## # i 77 more rows
```

*# This accomplishes the same thing using more pipes*

```
starwars.cut %>% group_by(gender) %>% arrange(desc(height), .by_group=TRUE)
```

```
## # A tibble: 87 x 9
## # Groups:   gender [3]
##   name      height  mass hair_color skin_color eye_color sex  gender homeworld
##   <chr>    <int> <dbl> <chr>    <chr>    <chr>  <chr> <chr> <chr>
## 1 Taun We      213  NA   none     grey     black  fema~ femin~ Kamino
## 2 Adi Gall~    184  50   none     dark     blue   fema~ femin~ Coruscant
## 3 Ayla Sec~    178  55   none     blue     hazel  fema~ femin~ Ryloth
## 4 Shaak Ti     178  57   none     red, blue~ black  fema~ femin~ Shili
## 5 Luminara~   170  56.2 black   yellow   blue   fema~ femin~ Mirial
## 6 Zam Wese~   168  55   blonde  fair, gre~ yellow  fema~ femin~ Zolan
## 7 Jocasta ~   167  NA   white    fair     blue   fema~ femin~ Coruscant
## 8 Barriss ~   166  50   black    yellow   blue   fema~ femin~ Mirial
## 9 Beru Whi~   165  75   brown    light    blue   fema~ femin~ Tatooine
## 10 Dormé      165  NA   brown    light    brown  fema~ femin~ Naboo
## # i 77 more rows
```

## 2. Subset a data frame using the filter function

```
head(starwars.cut)
```

```
## # A tibble: 6 x 9
##   name      height  mass hair_color skin_color eye_color sex  gender homeworld
##   <chr>      <int> <dbl> <chr>      <chr>      <chr>  <chr> <chr> <chr>
## 1 Luke Skyw~   172    77 blond      fair        blue   male  mascu~ Tatooine
## 2 C-3PO       167    75 <NA>      gold        yellow none  mascu~ Tatooine
## 3 R2-D2       96     32 <NA>      white, bl~ red     none  mascu~ Naboo
## 4 Darth Vad~  202   136 none       white       yellow male  mascu~ Tatooine
## 5 Leia Orga~  150    49 brown     light       brown  fema~ femin~ Alderaan
## 6 Owen Lars   178   120 brown, gr~ light       blue   male  mascu~ Tatooine
```

*# note there is a filter function in the stats package too*

```
starwars.cut %>% filter(homeworld=="Naboo")
```

```
## # A tibble: 11 x 9
##   name      height  mass hair_color skin_color eye_color sex  gender homeworld
##   <chr>      <int> <dbl> <chr>      <chr>      <chr>  <chr> <chr> <chr>
## 1 R2-D2       96     32 <NA>      white, bl~ red     none  mascu~ Naboo
## 2 Palpatine   170    75 grey       pale        yellow  male  mascu~ Naboo
## 3 Jar Jar ~   196    66 none       orange     orange  male  mascu~ Naboo
## 4 Roos Tar~   224    82 none       grey       orange  male  mascu~ Naboo
## 5 Rugor Na~   206    NA none       green      orange  male  mascu~ Naboo
## 6 Ric Olié    183    NA brown     fair        blue   <NA> <NA> Naboo
## 7 Quarsh P~   183    NA black     dark        brown  <NA> <NA> Naboo
## 8 Gregar T~   185    85 black     dark        brown  male  mascu~ Naboo
## 9 Cordé      157    NA brown     light       brown  fema~ femin~ Naboo
## 10 Dormé     165    NA brown     light       brown  fema~ femin~ Naboo
## 11 Padmé Am~  165    45 brown     light       brown  fema~ femin~ Naboo
```

```
filter(starwars.cut, homeworld=='Naboo' & hair_color != 'grey')
```

```
## # A tibble: 9 x 9
##   name      height  mass hair_color skin_color eye_color sex  gender homeworld
##   <chr>      <int> <dbl> <chr>      <chr>      <chr>  <chr> <chr> <chr>
## 1 Jar Jar B~   196    66 none       orange     orange  male  mascu~ Naboo
## 2 Roos Tarp~   224    82 none       grey       orange  male  mascu~ Naboo
## 3 Rugor Nass   206    NA none       green      orange  male  mascu~ Naboo
## 4 Ric Olié    183    NA brown     fair        blue   <NA> <NA> Naboo
## 5 Quarsh Pa~   183    NA black     dark        brown  <NA> <NA> Naboo
## 6 Gregar Ty~   185    85 black     dark        brown  male  mascu~ Naboo
## 7 Cordé      157    NA brown     light       brown  fema~ femin~ Naboo
## 8 Dormé     165    NA brown     light       brown  fema~ femin~ Naboo
## 9 Padmé Ami~  165    45 brown     light       brown  fema~ femin~ Naboo
```

```
starwars.cut %>% filter(height > 200) %>% filter(mass != 'NA')
```

```
## # A tibble: 7 x 9
##   name      height  mass hair_color skin_color eye_color sex  gender homeworld
```



```
##   <chr>      <int> <dbl> <chr>      <chr>      <chr>      <chr> <chr> <chr>
## 1 Darth Vad~    202   136 none      white      yellow     male  mascu~ Tatooine
## 2 Chewbacca    228   112 brown    unknown    blue       male  mascu~ Kashyyyk
## 3 Roos Tarp~   224    82 none      grey       orange     male  mascu~ Naboo
## 4 Lama Su      229    88 none      grey       black      male  mascu~ Kamino
## 5 Grievous     216   159 none      brown, wh~ green, y~  male  mascu~ Kalee
## 6 Tarfful      234   136 brown    brown      blue       male  mascu~ Kashyyyk
## 7 Tion Medon   206    80 none      grey       black      male  mascu~ Utapau
```

### 3. Group observations using the group\_by function

```
group_by(starwars.cut, gender)
```

```
## # A tibble: 87 x 9
## # Groups:   gender [3]
##   name      height  mass hair_color skin_color eye_color sex  gender homeworld
##   <chr>    <int> <dbl> <chr>      <chr>      <chr>      <chr> <chr> <chr>
## 1 Luke Sky~   172    77 blond     fair        blue       male  mascu~ Tatooine
## 2 C-3P0      167    75 <NA>      gold        yellow     none  mascu~ Tatooine
## 3 R2-D2      96     32 <NA>      white, bl~ red        none  mascu~ Naboo
## 4 Darth Va~  202   136 none      white       yellow     male  mascu~ Tatooine
## 5 Leia Org~  150    49 brown     light       brown      fema~ femin~ Alderaan
## 6 Owen Lars  178   120 brown, gr~ light       blue       male  mascu~ Tatooine
## 7 Beru Whi~  165    75 brown     light       blue       fema~ femin~ Tatooine
## 8 R5-D4      97     32 <NA>      white, red red        none  mascu~ Tatooine
## 9 Biggs Da~  183    84 black     light       brown      male  mascu~ Tatooine
## 10 Obi-Wan ~  182    77 auburn, w~ fair        blue-gray male  mascu~ Stewjon
## # i 77 more rows
```

```
ungroup(starwars.cut)
```

```
## # A tibble: 87 x 9
##   name      height  mass hair_color skin_color eye_color sex  gender homeworld
##   <chr>    <int> <dbl> <chr>      <chr>      <chr>      <chr> <chr> <chr>
## 1 Luke Sky~   172    77 blond     fair        blue       male  mascu~ Tatooine
## 2 C-3P0      167    75 <NA>      gold        yellow     none  mascu~ Tatooine
## 3 R2-D2      96     32 <NA>      white, bl~ red        none  mascu~ Naboo
## 4 Darth Va~  202   136 none      white       yellow     male  mascu~ Tatooine
## 5 Leia Org~  150    49 brown     light       brown      fema~ femin~ Alderaan
## 6 Owen Lars  178   120 brown, gr~ light       blue       male  mascu~ Tatooine
## 7 Beru Whi~  165    75 brown     light       blue       fema~ femin~ Tatooine
## 8 R5-D4      97     32 <NA>      white, red red        none  mascu~ Tatooine
## 9 Biggs Da~  183    84 black     light       brown      male  mascu~ Tatooine
## 10 Obi-Wan ~  182    77 auburn, w~ fair        blue-gray male  mascu~ Stewjon
## # i 77 more rows
```

### 4. Summarize your data frame with the summarize function

```
# average height over all observations
starwars.cut %>%
  mutate(height.feet=height/30.48) %>%
  summarize(mean=mean(height.feet,na.rm=TRUE))
```

```
## # A tibble: 1 x 1
##   mean
##   <dbl>
## 1  5.72
```

```
# average height by gender
starwars.cut %>%
  mutate(height.feet=height/30.48) %>%
  group_by(gender) %>%
  summarize(mean=mean(height.feet,na.rm=TRUE))
```

```
## # A tibble: 3 x 2
##   gender    mean
##   <chr>    <dbl>
## 1 feminine  5.40
## 2 masculine 5.79
## 3 <NA>      5.95
```

Now we will compute the deviation of each individual height (in feet) from the average for their gender

```
starwars.cut %>%
  mutate(height.feet=height/30.48) %>%
  group_by(gender) %>%
  mutate(height.dev=height.feet-mean(height.feet,na.rm=TRUE))%>%
  mutate(height.grp.mean=mean(height.feet,na.rm=TRUE))%>%
  select(name,gender,height.grp.mean,height.dev)
```

```
## # A tibble: 87 x 4
## # Groups:   gender [3]
##   name          gender height.grp.mean height.dev
##   <chr>         <chr>         <dbl>      <dbl>
## 1 Luke Skywalker masculine      5.79     -0.148
## 2 C-3P0         masculine      5.79     -0.312
## 3 R2-D2         masculine      5.79     -2.64
## 4 Darth Vader  masculine      5.79      0.836
## 5 Leia Organa  feminine      5.40     -0.482
## 6 Owen Lars    masculine      5.79      0.0487
## 7 Beru Whitesun lars feminine      5.40      0.0103
## 8 R5-D4        masculine      5.79     -2.61
## 9 Biggs Darklighter masculine      5.79      0.213
## 10 Obi-Wan Kenobi masculine      5.79      0.180
## # i 77 more rows
```

Now adding the deviation from the overall mean height

```

starwars.cut %>%
  mutate(height.feet=height/30.48) %>%
  group_by(gender) %>%
    mutate(height.dev=height.feet - mean(height.feet,na.rm=TRUE))%>%
    mutate(height.grp.mean=mean(height.feet,na.rm=TRUE)) %>%
  ungroup(gender) %>%
  mutate(height.overall.dev=height.feet - mean(height.feet,na.rm=TRUE),mean.overall=mean(hei
  select(name, gender, height.grp.mean, height.dev, mean.overall,height.overall.dev)

```

```

## # A tibble: 87 x 6
##   name      gender height.grp.mean height.dev mean.overall height.overall.dev
##   <chr>    <chr>         <dbl>     <dbl>         <dbl>         <dbl>
## 1 Luke Skywa~ mascul~           5.79     -0.148           5.72         -0.0774
## 2 C-3P0     mascul~           5.79     -0.312           5.72         -0.241
## 3 R2-D2     mascul~           5.79     -2.64            5.72         -2.57
## 4 Darth Vader mascul~           5.79      0.836           5.72          0.907
## 5 Leia Organa femin~           5.40     -0.482           5.72         -0.799
## 6 Owen Lars mascul~           5.79      0.0487          5.72          0.119
## 7 Beru White~ femin~           5.40      0.0103          5.72         -0.307
## 8 R5-D4     mascul~           5.79     -2.61            5.72         -2.54
## 9 Biggs Dark~ mascul~           5.79      0.213           5.72          0.284
## 10 Obi-Wan Ke~ mascul~           5.79      0.180           5.72          0.251
## # i 77 more rows

```

Example: Select all gender="feminine" individuals with height in feet greater than the overall group average height

```

starwars.cut %>%
  filter(gender=="feminine") %>%
  mutate(height.feet=height/30.48) %>%
  mutate(height.mean=mean(height.feet,na.rm=TRUE)) %>%
  filter(height.feet > height.mean,na.rm=TRUE)%>%
  select(name,gender,height.feet)

```

```

## # A tibble: 11 x 3
##   name      gender height.feet
##   <chr>    <chr>         <dbl>
## 1 Beru Whitesun lars feminine           5.41
## 2 Ayla Secura    feminine           5.84
## 3 Adi Gallia    feminine           6.04
## 4 Luminara Unduli feminine           5.58
## 5 Barriss Offee feminine           5.45
## 6 Dormé         feminine           5.41
## 7 Zam Wesell    feminine           5.51
## 8 Taun We       feminine           6.99
## 9 Jocasta Nu    feminine           5.48
## 10 Shaak Ti     feminine           5.84
## 11 Padmé Amidala feminine           5.41

```