# dplyr Basics

dplyr is a grammar of data manipulation, providing a consistent set of verbs that help you solve the most common data manipulation challenges:

* mutate() adds new variables that are functions of existing variables
* select() picks variables based on their names.
* filter() picks cases based on their values.
* summarise() reduces multiple values down to a single summary.
* arrange() changes the ordering of the rows.

## Loading Packages and Data

library(tidyverse)

## -- Attaching packages ------------------------------------ tidyverse 1.3.0 --

## v ggplot2 3.3.0 v purrr 0.3.3  
## v tibble 2.1.3 v dplyr 0.8.5  
## v tidyr 1.0.2 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.5.0

## -- Conflicts --------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

data("starwars")  
help("starwars")

## starting httpd help server ...

## done

## Filter

filters information based on a value (remember to use ==)

filter(starwars,films=="The Force Awakens")

## # A tibble: 5 x 13  
## name height mass hair\_color skin\_color eye\_color birth\_year gender homeworld  
## <chr> <int> <dbl> <chr> <chr> <chr> <dbl> <chr> <chr>   
## 1 Finn NA NA black dark dark NA male <NA>   
## 2 Rey NA NA brown light hazel NA female <NA>   
## 3 Poe ~ NA NA brown light brown NA male <NA>   
## 4 BB8 NA NA none none black NA none <NA>   
## 5 Capt~ NA NA unknown unknown unknown NA female <NA>   
## # ... with 4 more variables: species <chr>, films <list>, vehicles <list>,  
## # starships <list>

ForceAwakens <- filter(starwars,films=="The Force Awakens")

## Arrange

changes the ordering of the rows

arrange(ForceAwakens,name)

## # A tibble: 5 x 13  
## name height mass hair\_color skin\_color eye\_color birth\_year gender homeworld  
## <chr> <int> <dbl> <chr> <chr> <chr> <dbl> <chr> <chr>   
## 1 BB8 NA NA none none black NA none <NA>   
## 2 Capt~ NA NA unknown unknown unknown NA female <NA>   
## 3 Finn NA NA black dark dark NA male <NA>   
## 4 Poe ~ NA NA brown light brown NA male <NA>   
## 5 Rey NA NA brown light hazel NA female <NA>   
## # ... with 4 more variables: species <chr>, films <list>, vehicles <list>,  
## # starships <list>

ForceAwakens <- arrange(ForceAwakens,name)

## Mutate

adds new variables that are functions of existing variables

Droid <- filter(starwars,species=="Droid")  
  
mutate(Droid,mass\_lbs = mass\*2.2)

## # A tibble: 5 x 14  
## name height mass hair\_color skin\_color eye\_color birth\_year gender homeworld  
## <chr> <int> <dbl> <chr> <chr> <chr> <dbl> <chr> <chr>   
## 1 C-3PO 167 75 <NA> gold yellow 112 <NA> Tatooine   
## 2 R2-D2 96 32 <NA> white, bl~ red 33 <NA> Naboo   
## 3 R5-D4 97 32 <NA> white, red red NA <NA> Tatooine   
## 4 IG-88 200 140 none metal red 15 none <NA>   
## 5 BB8 NA NA none none black NA none <NA>   
## # ... with 5 more variables: species <chr>, films <list>, vehicles <list>,  
## # starships <list>, mass\_lbs <dbl>

Droid <- mutate(Droid,mass\_lbs = mass\*2.2)

# Select

picks variables based on their names

Droid2 <- select(Droid,name,height,mass\_lbs,homeworld)  
Droid3 <- select(Droid,name,height,mass\_lbs,homeworld,everything())

# Summarize

reduces multiple values down to a single summary

summarize(starwars,mass=mean(mass,na.rm=TRUE))

## # A tibble: 1 x 1  
## mass  
## <dbl>  
## 1 97.3

starwars\_summary <- group\_by(starwars,species)  
summarize(starwars\_summary,mass=mean(mass,na.rm=TRUE))

## # A tibble: 38 x 2  
## species mass  
## <chr> <dbl>  
## 1 Aleena 15   
## 2 Besalisk 102   
## 3 Cerean 82   
## 4 Chagrian NaN   
## 5 Clawdite 55   
## 6 Droid 69.8  
## 7 Dug 40   
## 8 Ewok 20   
## 9 Geonosian 80   
## 10 Gungan 74   
## # ... with 28 more rows