Third assignment: solutions

Exercise

Use the starwars data set in the dplyr package to:

•list the different human characters,
•list the different worlds,
•compute the average weight and height of the different character types,
•display on a plot the number of characters of each type in a deacresing order,

•visualize the relationship between the height and weight of the different characters.

list the different human characters

```
?starwars
starwars$name[starwars$species == 'Human']
```

```
## [1] "Luke Skywalker" "Darth Vader" "Leia Organa »
 ## [4] "Owen Lars" "Beru Whitesun lars" "Biggs Darklighter"
## [7] "Obi-Wan Kenobi" "Anakin Skywalker" "Wilhuff Tarkin"
## [10] "Han Solo" "Wedge Antilles" "Jek Tono Porkins"
## [13] "Palpatine" "Boba Fett" "Lando Calrissian"
## [16] "Lobot" "Mon Mothma" "Arvel Crynyd"
## [19] "Qui-Gon Jinn" "Finis Valorum" NA
## [22] NA "Shmi Skywalker" "Mace Windu"
## [25] "Gregar Typho" "Cordé" "Cliegg Lars"
## [28] "Dormé" "Dooku" "Bail Prestor Organa »
 ## [31] "Jango Fett" "Jocasta Nu" NA
## [34] "Raymus Antilles" NA "Finn"
## [37] "Rey" "Poe Dameron" NA
## [40] "Padmé Amidala"
```

list the different worlds

```
unique(starwars$homeworld)
```

```
## [1] "Tatooine" "Naboo" "Alderaan" "Stewjon"
## [5] "Eriadu" "Kashyyyk" "Corellia" "Rodia »
 ## [9] "Nal Hutta" "Bestine IV" NA "Kamino"
## [13] "Trandosha" "Socorro" "Bespin" "Mon Cala"
## [17] "Chandrila" "Endor" "Sullust" "Cato Neimoidia"
## [21] "Coruscant" "Toydaria" "Malastare" "Dathomir"
## [25] "Ryloth" "Vulpter" "Troiken" "Tund"
## [29] "Haruun Kal" "Cerea" "Glee Anselm" "Iridonia"
## [33] "Iktotch" "Quermia" "Dorin" "Champala"
## [37] "Geonosis" "Mirial" "Serenno" "Concord Dawn"
## [41] "Zolan" "Ojom" "Aleen Minor" "Skako"
## [45] "Muunilinst" "Shili" "Kalee" "Umbara"
## [49] "Utapau"
```

compute the average weight and height of the different character types

```
weight = tapply(starwars$mass, starwars$species, mean)
height = tapply(starwars$height, starwars$species, mean)
Species = data.frame(weight,height)
library(ggplot2)

ggplot(data = Species) + geom_point(aes(x = weight,y =
height)) + geom_text(aes(x = weight,y = height,label =
rownames(Species)),nudge_y = 10)
```
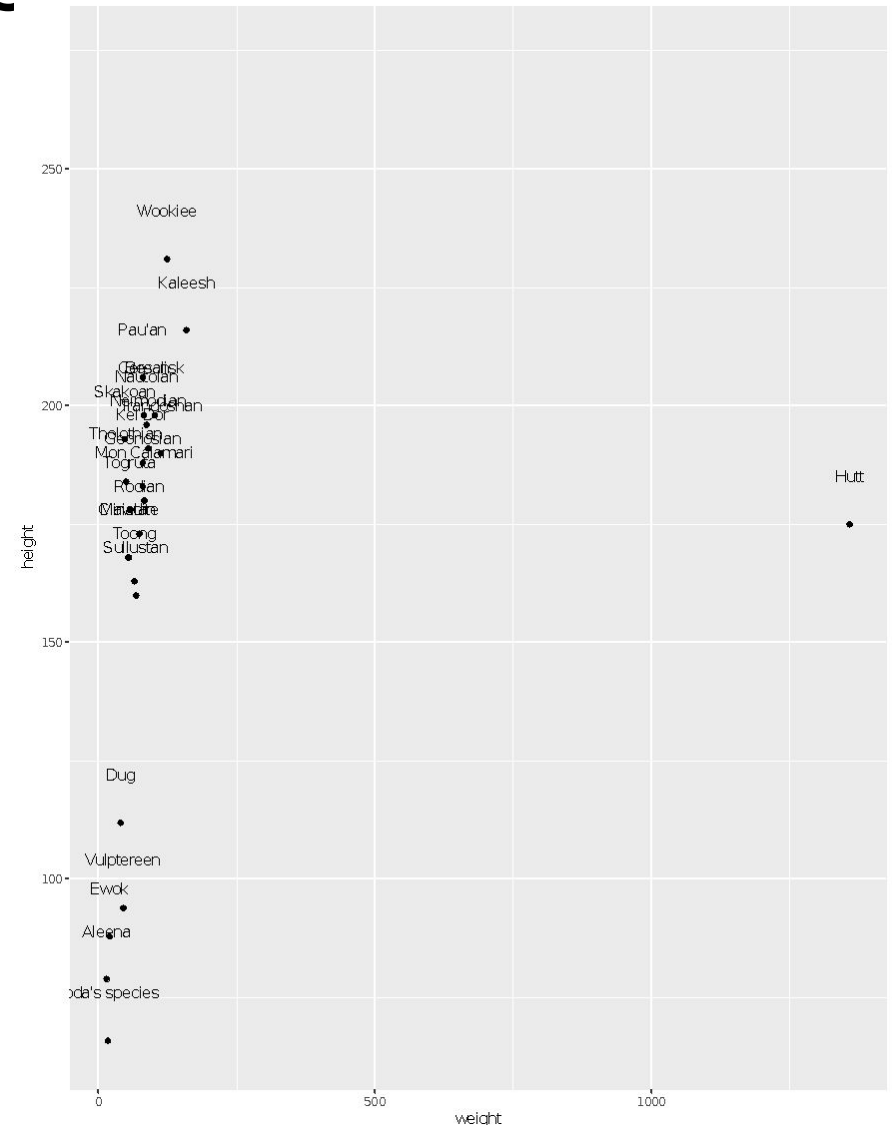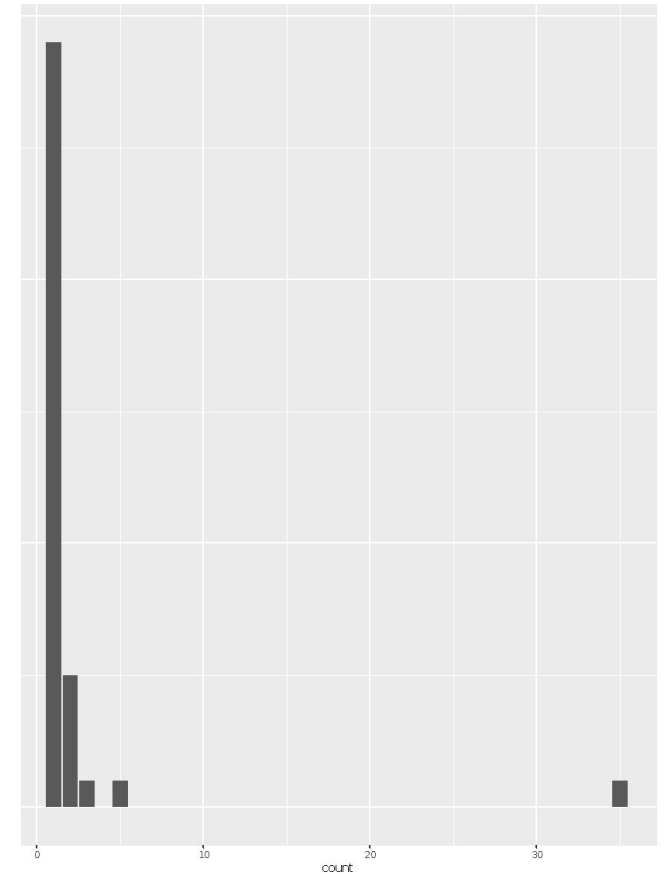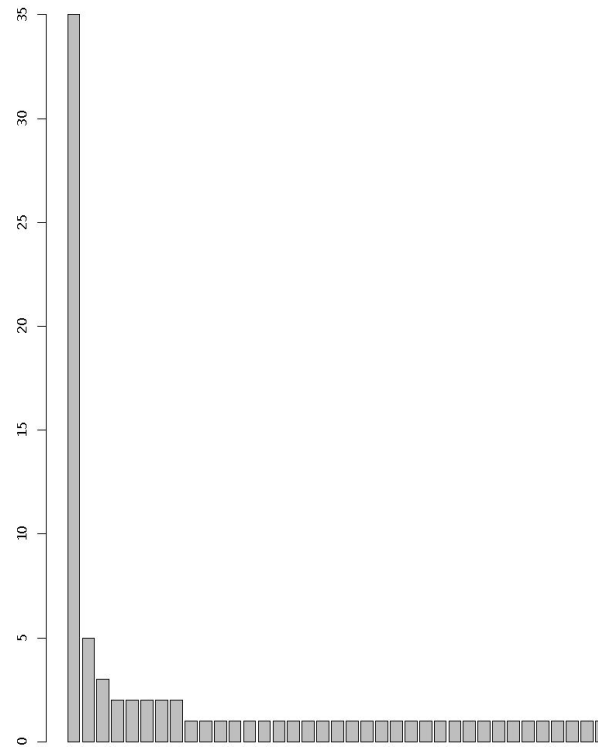
# display on a plot the number of characters of each type in a deacresing order

```
Species$count = table(starwars$species)
Species = Species[order(Species$count,decreasing = TRUE),]

ggplot(data = Species) + geom_bar(aes(x = count))
barplot(Species$count)
```

visualize the relationship between the height and weight of the different characters

```
ggplot(data = starwars) + geom_point(aes(x = mass,y = height))
+ geom_smooth(aes(x = mass,y = height),method = 'lm')
```
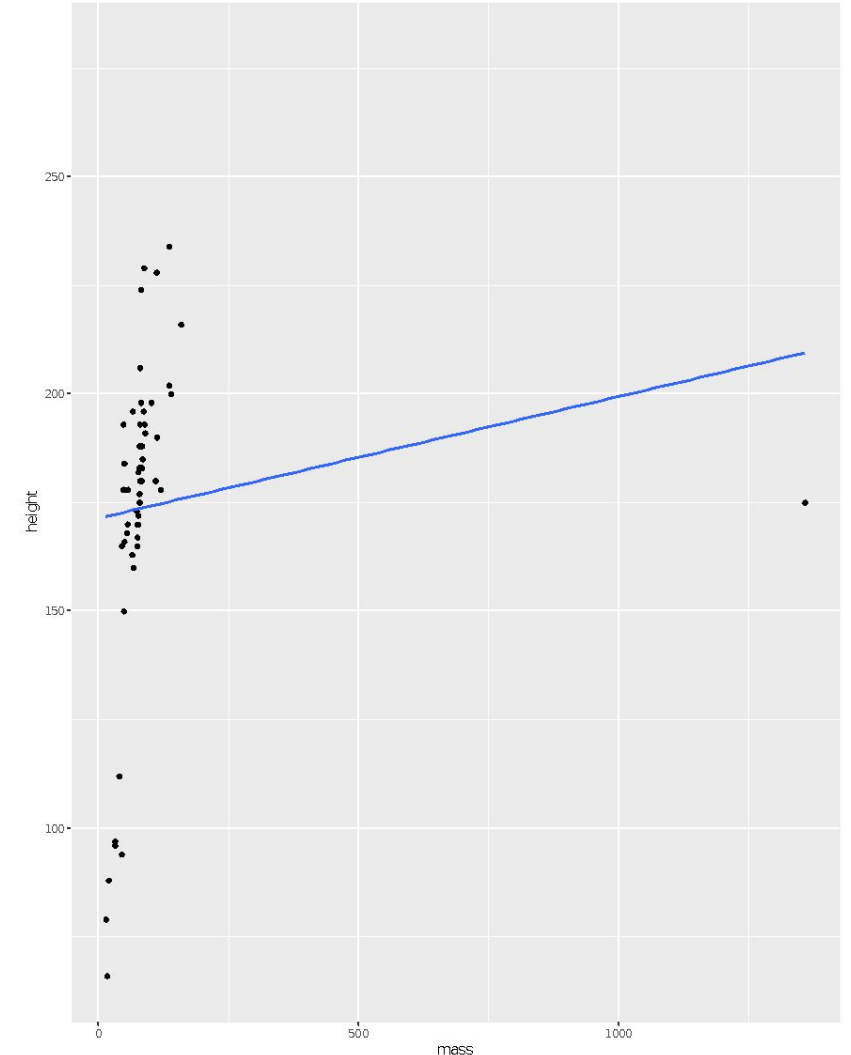
Exercise

Compare two simulated datasets with a plot and a hypothesis test.
Use a functions that:
 visualises the two distributions with a histogram,
 uses a t-test to compares means and summarises the results with a string

Result

```r
library(ggplot2)

histogram <- function(x1, x2, binwidth = 0.1, xlim = c(-3, 3)) {
  df <- data.frame(
    x = c(x1, x2),
    g = c(rep("x1", length(x1)), rep("x2", length(x2)))
  )

  ggplot(df, aes(x, fill = g)) +
    geom_histogram(binwidth = binwidth) +
    coord_cartesian(xlim = xlim)
}

t_test <- function(x1, x2) {
  test <- t.test(x1, x2)

  sprintf(
    "p value: %0.3f\n[%0.2f, %0.2f]",
    test$p.value, test$conf.int[1], test$conf.int[2]
  )
}
```

Result

```
x1 <- rnorm(100, mean = 0, sd = 0.5)
x2 <- rnorm(200, mean = 0.15, sd = 0.9)

histogram(x1, x2)
cat(t_test(x1, x2))
#> p value: 0.131
#> [-0.30, 0.04]
```