

# Dplyr.R

vjvic

2021-08-19

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

download.file("https://doi.org/10.1371/journal.pone.0149098.s002", "df.csv")
df = read.csv("df.csv")
unique(df$Prey.species)

##   [1] "Kudu"                "Unknown"
##   [3] ""                   "Hartebeest"
##   [5] "Warthog"             "kudu"
##   [7] "Ostrich"             "Zebra"
##   [9] "Buffalo"            "Eland"
##  [11] "Small Antelope"     "Bushbuck"
##  [13] "Medium Antelope"   "Hartebeest+Ostrich"
##  [15] "Ostrich, Small Antelope" "Hartebest"
##  [17] "Grysbok x2"        "Ostrich X2"
##  [19] "Porcupine"         "Ostrich, Eland"
##  [21] "Busbuck"           "Elephant"
##  [23] "Bushpig"           "Antelope (Small)"
##  [25] "Ostrich, Warthog"   "Hartbeest"
##  [27] "Warthog+Ostrich"    "Ostrich+Kudu"
##  [29] "Ostrich "           "Hartebeestx2"
##  [31] "Buffalo x2"        "Small Ant"

names(df)

##   [1] "GPS.Cluster"
##   [2] "Lion.ID"
##   [3] "Lion.sex"
##   [4] "State..kill..1..or.non.kill..0.."
##   [5] "Prey.species"
##   [6] "Average.wind.speed.precedding.GPS.cluster"
##   [7] "Moon.present..1..or.absent..0.."
##   [8] "Viewshed.50m"
##   [9] "Viewshed.50.to.100m"
##  [10] "Viewshed.100.to.300m"
##  [11] "Viewshed.300m"
##  [12] "Distance.to.downwind.cover"
##  [13] "Minimum.distance.to.cover"

df_1 = filter(df, Prey.species != "Unknown" & nchar(Prey.species)>0)%>%
  mutate(Prey.species = tolower(Prey.species))%>%
  group_by(Lion.ID, Lion.sex, Prey.species)%>%
  summarise(nb_killed = sum(State..kill..1..or.non.kill..0..))%>%
  arrange(desc(nb_killed))%>%
```

```

select(Lion = Lion.ID, prey = Prey.species, nb_killed)

## `summarise()` has grouped output by 'Lion.ID', 'Lion.sex'. You can override using the `.groups`
argument.

## Adding missing grouping variables: `Lion.sex`

df_1

## # A tibble: 67 x 4
## # Groups:   Lion, Lion.sex [6]
##   Lion.sex Lion    prey      nb_killed
##   <chr>    <chr>  <chr>      <int>
## 1 Female   Jess    ostrich     38
## 2 Male     John    ostrich     33
## 3 Female   Gina     kudu       22
## 4 Female   Aardlam  kudu       20
## 5 Female   Josie    kudu       16
## 6 Female   Aardlam  warthog     14
## 7 Female   Jess     hartebeest  14
## 8 Male     John     eland       14
## 9 Male     John     hartebeest  14
## 10 Male    John     zebra       12
## # ... with 57 more rows

attach(df_1)

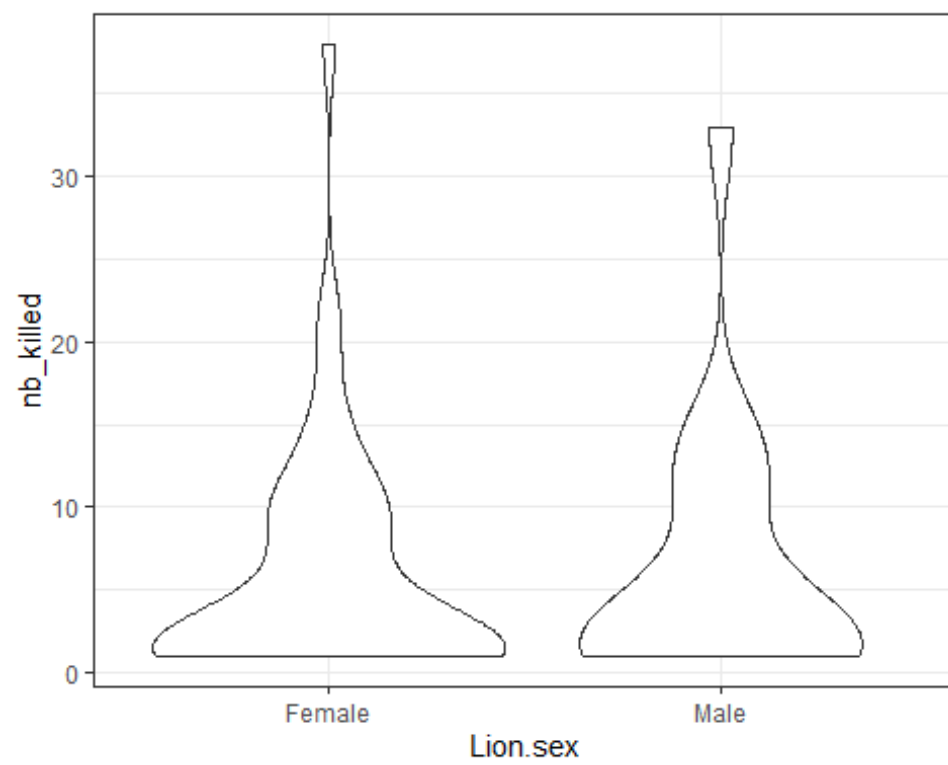
library(tidyr)
df_spread = spread(df_1, Lion, nb_killed)
df_spread

## # A tibble: 38 x 8
## # Groups:   Lion.sex [2]
##   Lion.sex prey      Aardlam  Gina  Jess  John  Josie  Roy
##   <chr>    <chr>      <int> <int> <int> <int> <int> <int>
## 1 Female  antelope (small)    NA    NA    1    NA    NA    NA
## 2 Female  buffalo            2     3    2    NA    7    NA
## 3 Female  buffalo x2         NA    NA    NA    NA    1    NA
## 4 Female  busbuck           NA    NA    1    NA    NA    NA
## 5 Female  bushbuck          NA     2    NA    NA    NA    NA
## 6 Female  bushpig           NA    NA    1    NA    NA    NA
## 7 Female  eland             1     9   11    NA   11    NA
## 8 Female  elephant          NA    NA    1    NA    2    NA
## 9 Female  grysbok x2        NA    NA    1    NA    NA    NA
## 10 Female  hartbeest         NA    NA    1    NA    NA    NA
## # ... with 28 more rows

library(ggplot2)

ggplot(df_1, aes(Lion.sex, nb_killed)) +
  geom_violin() + theme_bw()

```



```
wilcox.test(nb_killed[Lion.sex == "Male"],
            nb_killed[Lion.sex == "Female"])

## Warning in wilcox.test.default(nb_killed[Lion.sex == "Male"], nb_killed[Lion.sex
## == : cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: nb_killed[Lion.sex == "Male"] and nb_killed[Lion.sex == "Female"]
## W = 460.5, p-value = 0.7789
## alternative hypothesis: true location shift is not equal to 0
```