

# Data manipulation with pipes

Charles

2023-03-07

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   :  2.00
##  1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##  Mean   :15.4    Mean   : 42.98
##  3rd Qu.:19.0    3rd Qu.: 56.00
##  Max.   :25.0    Max.   :120.00
```

## Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

## Sequential Code

```
numbers <- -300:546
numbers <- 1:300
numbers_mean <- mean(numbers)
sqrt(x = numbers_mean)
```

```
## [1] 12.26784
```

## Nested Code

```
numbers <- 1:300
mean(numbers)
```

```
## [1] 150.5
```

```
sqrt(mean(numbers))
```

```
## [1] 12.26784
```

## Piping code

implement with 'magrittr'

```
library(magrittr)
```

pipe symbol is %>% or |

```
1:300 %>% mean() %>% sqrt()
```

```
## [1] 12.26784
```

```
surveys <- read.csv(file = "../data-raw/surveys.csv")
str(surveys)
```

```
## 'data.frame': 35549 obs. of 9 variables:
## $ record_id : int 1 2 3 4 5 6 7 8 9 10 ...
## $ month : int 7 7 7 7 7 7 7 7 7 7 ...
## $ day : int 16 16 16 16 16 16 16 16 16 16 ...
## $ year : int 1977 1977 1977 1977 1977 1977 1977 1977 1977 1977 ...
## $ plot_id : int 2 3 2 7 3 1 2 1 1 6 ...
## $ species_id : chr "NL" "NL" "DM" "DM" ...
## $ sex : chr "M" "M" "F" "M" ...
## $ hindfoot_length: int 32 33 37 36 35 14 NA 37 34 20 ...
## $ weight : int NA NA NA NA NA NA NA NA NA NA ...
```

```
surveys$year %>% mean()
```

```
## [1] 1990.475
```

```
surveys$weight %>% mean(na.rm = TRUE)
```

```
## [1] 42.67243
```

```
surveys1 <- select(surveys, year, month, day, species_id)
```

```
surveys2 <- select(surveys, year, species_id, weight) %>% filter(!is.na(weight)) %>%
mutate(weight_kg = weight/1000) %>% select(year, species_id, weight_kg) %>%
filter(species_id == 'SH')
str(surveys2)
```

```
## 'data.frame': 141 obs. of 3 variables:
## $ year : int 1978 1982 1982 1986 1987 1987 1987 1987 1987 1988 ...
## $ species_id: chr "SH" "SH" "SH" "SH" ...
## $ weight_kg : num 0.089 0.106 0.052 0.055 0.077 0.078 0.104 0.058 0.052 0.06 ...
```

```
surveys2_filtered <- filter(surveys2, species_id == 'SH')
str(surveys2_filtered)
```

```
## 'data.frame': 141 obs. of 3 variables:
## $ year : int 1978 1982 1982 1986 1987 1987 1987 1987 1987 1988 ...
## $ species_id: chr "SH" "SH" "SH" "SH" ...
## $ weight_kg : num 0.089 0.106 0.052 0.055 0.077 0.078 0.104 0.058 0.052 0.06 ...
```