Tidyverse - Solutions

Exercises

- Install the Tidyverse or update it. Then load it.

```
#install.packages("tidyverse") #if not installed/updated yet
library(tidyverse)
```

- Load the `magrittr` package and check out its help file on the forward pipe operator (`%>%`).

```
library(magrittr)
?magrittr
```

- Let’s compare the Base R approach with the Tidyverse approach:
- First, install the package `palmerpenguins` and load the built-in penguins dataset into RStudio. You can do this by typing `data(package = "palmerpenguins")`. How many rows and variables does the `penguins` dataset have? And how many factors?

```
#install.packages("palmerpenguins") #if not installed yet

#load package
library(palmerpenguins)

#load the datasets
data(package = "palmerpenguins")

#dimensions of the penguins dataset
glimpse(penguins)
# --> 344 rows, 8 variables, 3 factors
```

- Now select the following variables: `species`, `island`, `flipper_length_mm`, `sex`, `year`, once with the Base R approach, and once with the Tidyverse approach

```
#base r
penguins[, c("species", "island", "flipper_length_mm", "sex", "year")]

#tidyverse
penguins %>%
select(species, island, flipper_length_mm, sex, year)
```

- Remove all NAs. Again, once with the Base R and once with the Tidyverse approach.

```
#base r
na.omit(penguins)

#tidyverse
penguins %>%
  drop_na()
```

- What's the median flipper length of the different species? Try to obtain the same results with both approaches.

```
#base r
penguins <- penguins[!(is.na(penguins$bill_length_mm)),]

tapply(penguins$bill_length_mm, penguins$species, median)
```

```
##    Adelie Chinstrap    Gentoo
##     38.80     49.55     47.30
```

```
#tidyverse
penguins %>%
  group_by(species) %>%
  summarise(median_flipper_length = median(bill_length_mm, na.rm = T))
```

```
## # A tibble: 3 x 2
##   species   median_flipper_length
##   <fct>                     <dbl>
## 1 Adelie                     38.8
## 2 Chinstrap                  49.6
## 3 Gentoo                     47.3
```

- What's the minimum body mass on the different islands? Again, try to find the same results with both approaches.

```
#base r
penguins <- penguins[!(is.na(penguins$body_mass_g)),]

tapply(penguins$body_mass_g, penguins$island, min)
```

```
##    Biscoe     Dream Torgersen
##      2850      2700      2900
```

```
#tidyverse
penguins %>%
  group_by(island) %>%
  summarise(min_bodymass = min(body_mass_g, na.rm = T))
```

```
## # A tibble: 3 x 2
##   island    min_bodymass
##   <fct>            <int>
## 1 Biscoe            2850
## 2 Dream             2700
## 3 Torgersen         2900
```