Janik_BT4BR_ex7

Using of tidyr and dplyr libraries

1. Loading the "iris" dataset and showing summary. Transforming dataset into long format changing columns connected to trait into one column named Trait and assigned values to each row are stored in Size column.

```
library(datasets)
library(tidyr)
data(iris)
summary(iris)
```

```
  Sepal.Length     Sepal.Width      Petal.Length     Petal.Width
 Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100
 1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
 Median :5.800   Median :3.000   Median :4.350   Median :1.300
 Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
 3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
 Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
       Species
 setosa    :50
 versicolor:50
 virginica :50
```

```
iris_longer <- iris %>%
  pivot_longer(cols = c(Sepal.Length,Sepal.Width,Petal.Length,Petal.Width), names_to = '
View(iris_longer)
```

2. Calculating mean length, width of the sepal and petal of each species.

```
library(dplyr)
```

```
Dołączanie pakietu: 'dplyr'

Następujące obiekty zostały zakryte z 'package:stats':

    filter, lag

Następujące obiekty zostały zakryte z 'package:base':

    intersect, setdiff, setequal, union
```

```
iris_longer_1 <- iris_longer %>%
  group_by(Species,Trait) %>%
  summarise(mean_trait = mean(Size))
```

```
`summarise()` has grouped output by 'Species'. You can override using the
`.groups` argument.
```

```
View(iris_longer_1)
```

I've used dplyr library in order to use group_by and summarise functions. Firstly I've ordered data by species and trait and then from this grouped data I took mean from column size.

3. **Finding the largest and widest, petal and sepal, of virginica and versicolor species.**

```
iris_longer_2 <- iris_longer %>%
  group_by(Species, Trait) %>%
  filter(Size == max(Size), Species %in% c("virginica","versicolor"))
View(iris_longer_2)
```

Firstly I've grouped data (using group_by from dplyr) into species and trait so I could take maximum value of grouped records. Then I've filtered only maximum values (it includes also equal maximum values) and species that was specified in task

4. **Calculating how many observations are above mean**

```
iris_longer_3 <- iris_longer %>%
  group_by(Species,Trait) %>%
  mutate(mean_trait = mean(Size)) %>%
  filter(Size > mean_trait) %>%
  summarise(records_above_mean = n())
```

```
`summarise()` has grouped output by 'Species'. You can override using the
`.groups` argument.
```

```
View(iris_longer_3)
```

Firstly I've grouped records by species and trait to calculate mean and opposing to what I did in iris_longer_1 I didn't use summarise because it would remove remaining columns and I wanted to use column size to filter results above mean. Then I've used summarise to count records above mean.

I've used LLM to check how to count number of records.