Using eurostat to obtain music related data

This document goes over how the `eurostat` R package can be used to retrieve music, and other cultural data.

Setup

To begin we need to install the packages needed in this document.

```
# Installing eurostat
install.packages("eurostat")
# Installing other packages used in this document
install.packages("dplyr")
install.packages("ggplot2")
install.packages("stringr")
```

After installing we can load the package to get access to the functions provided by it.

```
# Loading eurostat
library(eurostat)
# Other useful packages
library(dplyr) # helps with data manipulation
library(ggplot2) # helps with data vizualisation
```

Finding and retrieving the data

Let's start by going over how to search for datasets from the Eurostat database. This can be done by using function `search_eurostat()`.

```
# Searching for datasets
ym <- search_eurostat("cultural or sport")
```

We can now take a look at the datasets that match our search phrase.

```
# First six results from the search
head(ym)
```

```
# A tibble: 6 x 9
  title        code  type  last.update.of.data last.table.structure~1 data.start
  <chr>        <chr> <chr> <chr>               <chr>                  <chr>
1 Persons par~ ilc_~ data~ 18.07.2025          18.07.2025             2015
2 Persons par~ ilc_~ data~ 18.07.2025          18.07.2025             2015
3 Persons not~ ilc_~ data~ 18.07.2025          04.12.2024             2015
4 Persons par~ ilc_~ data~ 18.07.2025          23.04.2024             2022
5 Persons not~ ilc_~ data~ 18.07.2025          04.12.2024             2022
6 Persons par~ ilc_~ data~ 18.07.2025          22.04.2024             2006
# i abbreviated name: 1: last.table.structure.change
# i 3 more variables: data.end <chr>, values <dbl>, hierarchy <dbl>
```

We see that the there are several datasets about participation in cultural or sport activities. Let's take a close look at at the dataset "Persons participating in cultural or sport activities in the last 12 months by income quintile, household composition, degree of urbanisation and frequency." The code for this dataset is "ilc_scp04." After we have the dataset we can use the code corresponding to the dataset to obtain retrieve it from the Eurostat database. This can be done using the `get_eurostat()` function.

```
# Retrieving the dataset
y <- get_eurostat(id = "ilc_scp04")
```

Let's take a look at the dataset.

```
# First six rows of the dataset
head(y)
```

```
# A tibble: 6 x 10
  freq  frequenc acl00 deg_urb hhcomp quant_inc unit  geo   TIME_PERIOD values
  <chr> <chr>    <chr> <chr>   <chr>  <chr>     <chr> <chr> <date>       <dbl>
1 A     1-3      AC521 DEG1    A1     QU1       PC    AT    2015-01-01    29
2 A     1-3      AC521 DEG1    A1     QU1       PC    AT    2022-01-01    21.9
3 A     1-3      AC521 DEG1    A1     QU1       PC    BE    2015-01-01    18.1
4 A     1-3      AC521 DEG1    A1     QU1       PC    BE    2022-01-01    12.2
5 A     1-3      AC521 DEG1    A1     QU1       PC    BG    2015-01-01     8
6 A     1-3      AC521 DEG1    A1     QU1       PC    BG    2022-01-01     3.5
```

We can see that by default the dataset retrieved uses code for the column values. We can change this by using the `label_eurostat()` function.

```
# Labeling the codes in the dataset
y <- label_eurostat(y)
```

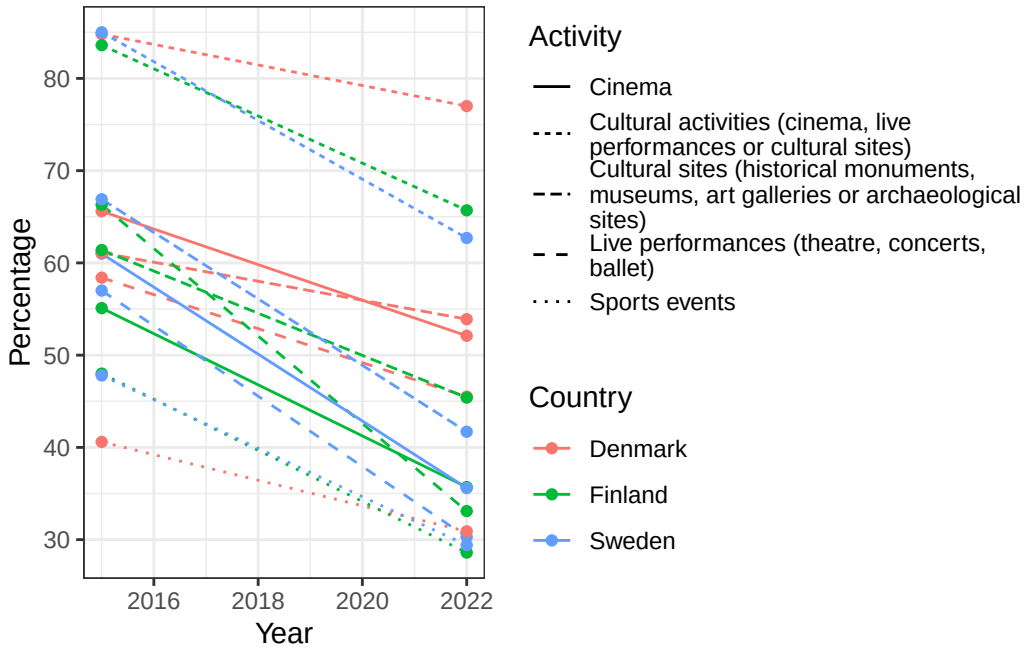
Let's look at the data again.

```
# First six rows of the dataset
head(y)
```

```
# A tibble: 6 x 10
  freq   frequenc  acl00 deg_urb hhcomp quant_inc unit  geo   TIME_PERIOD values
  <chr>  <chr>     <chr> <chr>   <chr>  <chr>     <chr> <chr> <date>       <dbl>
1 Annual From 1 t~ Cine~ Cities  One a~ First qu~ Perc~ Aust~ 2015-01-01    29
2 Annual From 1 t~ Cine~ Cities  One a~ First qu~ Perc~ Aust~ 2022-01-01    21.9
3 Annual From 1 t~ Cine~ Cities  One a~ First qu~ Perc~ Belg~ 2015-01-01    18.1
4 Annual From 1 t~ Cine~ Cities  One a~ First qu~ Perc~ Belg~ 2022-01-01    12.2
5 Annual From 1 t~ Cine~ Cities  One a~ First qu~ Perc~ Bulg~ 2015-01-01     8
6 Annual From 1 t~ Cine~ Cities  One a~ First qu~ Perc~ Bulg~ 2022-01-01     3.5
```

Now the dataset values are labelled. Let's plot the percentage of people who attended different kinds of cultural or sport activities at least once in Finland, Denmark and Sweden.

```
y %>%
    filter(geo %in% c("Finland", "Sweden", "Denmark"), deg_urb == "Total",
           frequenc == "At least once", hhcomp == "Total", quant_inc == "Total") %>%
    mutate(acl00 = stringr::str_wrap(acl00, width = 40)) %>%
    ggplot(aes(x = TIME_PERIOD, y = values, color = geo, linetype = acl00)) +
    geom_line() + geom_point() + theme_bw() +
    labs(color = "Country", y = "Percentage", x = "Year", linetype = "Activity")
```



The dataset has values at two time points 2015 and 2022. We can see that all the activities have seen decrease in attendance.

Conclusion

This document has provided an cursory overview of how the `eurostat` package can be used to retrieve music and culture related data from the Eurostat database. For more detailed tutorial of the package you can check out eurostat package's website or GitHub page. For more details about the Eurostat, you can check the database's website.