Assignment 2 - Data Visualization

Name: Penca Matei

Student number: s4039696

Assignment 2

Packages

```
library(ggplot2)
library(ggridges)
```

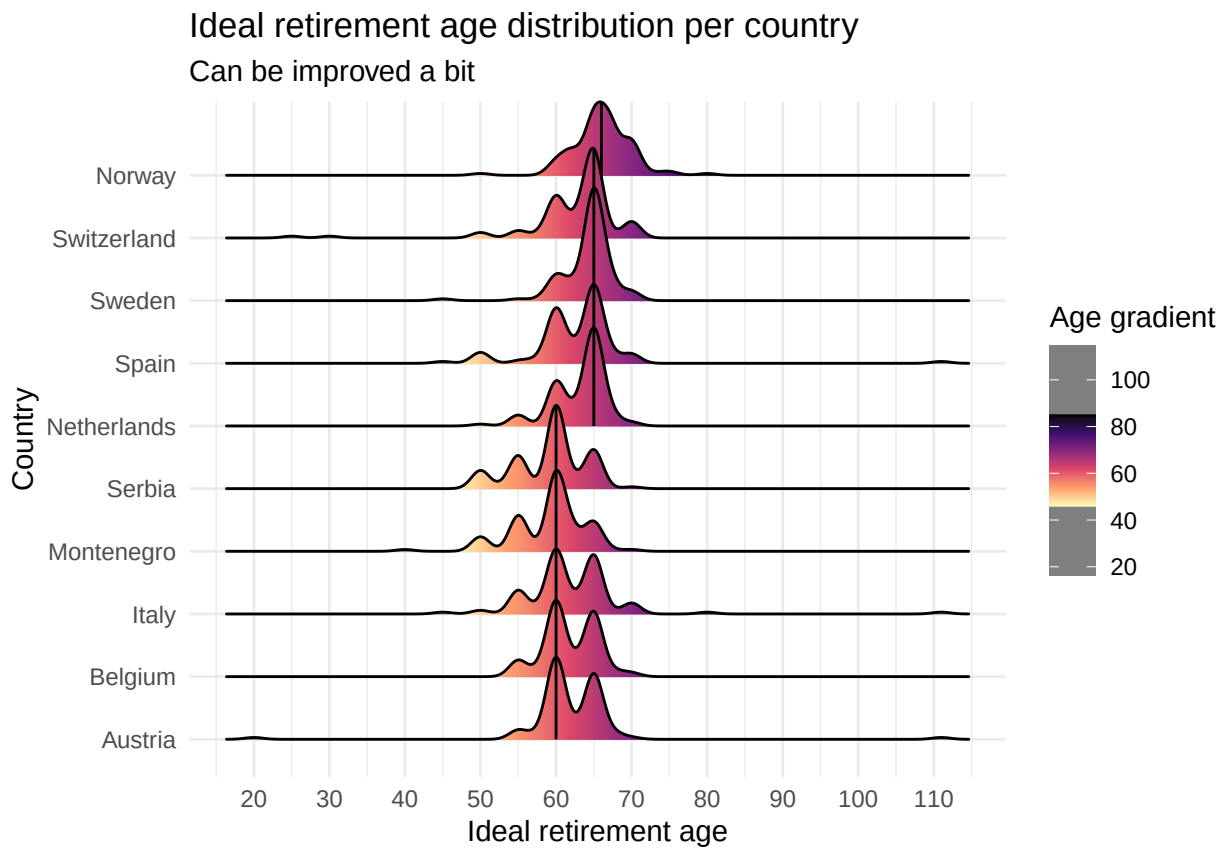
Data

```
data <- read.csv2("http://stulp.gmw.rug.nl/dataviz/ESS.csv", header = TRUE)
```

Code to produce visualization

First version of the code and graph:

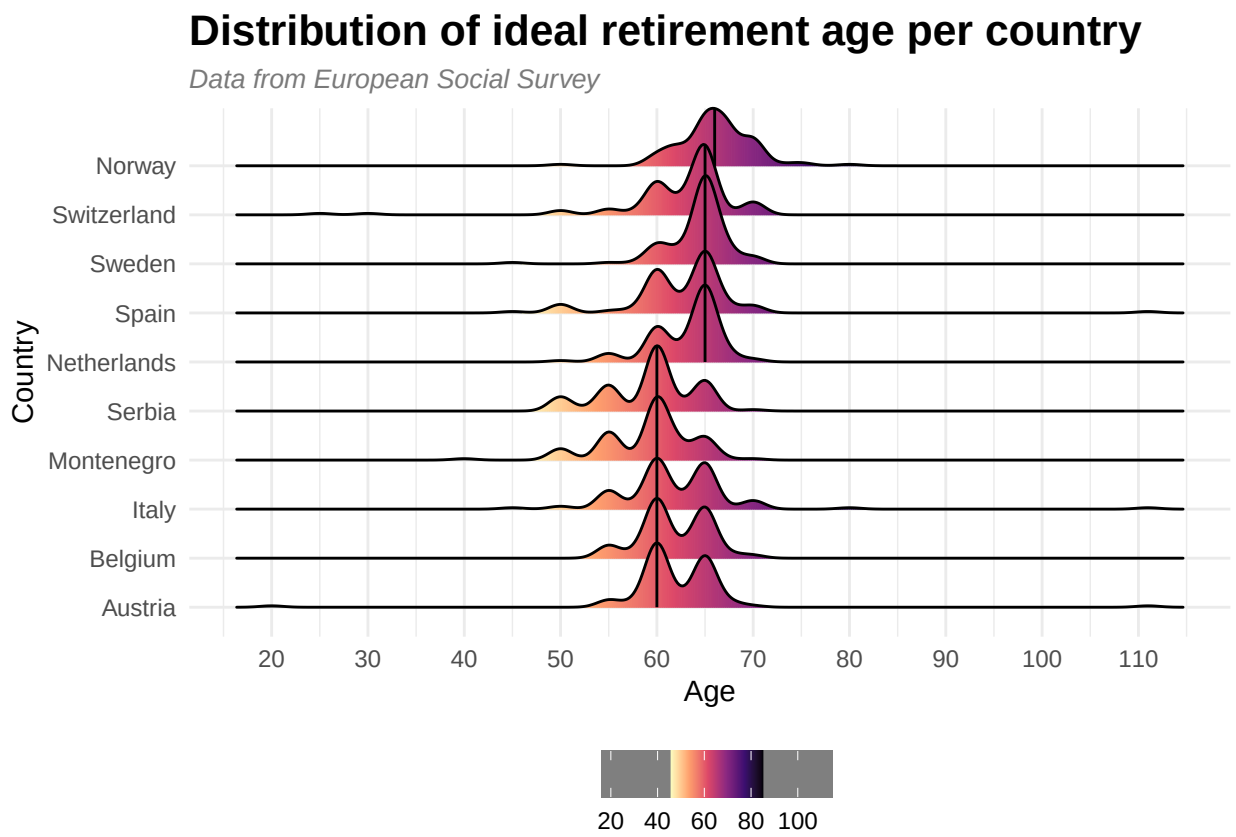
```
ggplot(data, aes(
  x = iagrtr, y = reorder(cntry, iagrtr, FUN = "median"), fill = ..x..
)) +
  geom_density_ridges_gradient(
    quantile_lines = TRUE, quantiles = 2
  ) +
  # Set the legend name, color, order and set the range for the gradient
  scale_fill_viridis_c(
    name = "Age gradient", option = "A", direction = -1, values = c(0.3, 0.7)
  ) +
  # In case any of the graphs get cut by the plane
  coord_cartesian(clip = "off") +
  scale_x_continuous(breaks = seq(20, 130, by = 10)) +
  labs(x = "Ideal retirement age", y = "Country",
       title = "Ideal retirement age distribution per country",
       subtitle = "Can be improved a bit") +
  theme_minimal()
```



We can still improve the look of the graph and we can also simplify it by adding some themes:

```
custom_theme <- theme(
  legend.position = "bottom",
  plot.title = element_text(size = 16, face = "bold", colour = "black"),
  plot.subtitle = element_text(size = 10, face = "italic", colour = "gray48")
)
```

```
ggplot(data, aes(
  x = iagrtr, y = reorder(cntry, iagrtr, FUN = "median"), fill = ..x..
)) +
  geom_density_ridges_gradient(
    quantile_lines = TRUE, quantiles = 2
  ) +
  # Set the legend name, color, order and set the range for the gradient
  scale_fill_viridis_c(
    name = "", option = "A", direction = -1, values = c(0.3, 0.7)
  ) +
  # In case any of the graphs get cut by the plane
  coord_cartesian(clip = "off") +
  scale_x_continuous(breaks = seq(20, 130, by = 10)) +
  labs(x = "Age", y = "Country",
       title = "Distribution of ideal retirement age per country",
       subtitle = "Data from European Social Survey") +
  # I combine the minimal theme with mine so that I don't rewrite code
  theme_minimal() +
  custom_theme
```



Description of visualization

Having sorted the graph by the median of the age, we can see that we have 5 countries with a median around 65 years old and 5 countries with one averaging 60 years old. We can see that northern countries have a higher retirement age. Another insight from this graph might be that in countries where the age is higher, the life expectancy might be longer and the stress people experience lower. Moreover, looking at the first countries, the distribution is uniform while for the bottom ones not so much(more ridges in the graph).

This trend can be observed if we look at the colors on the graph. In countries such as Norway or Sweden, the graph is predominantly purple(uniform distribution), but on the lower graphs, we can see it going from bright yellow to purple which also explains why the median is lower. As such a simple way of interpreting the graph would be: darker color means higher ideal retirement age while brighter color means the opposite.

It would have been nicer if I could also add points for the data entries, but because of the nature of the graph it won't look good and it will also load it too much. The graph also does not explicitly show the population size for each country and it is also unclear from the graph exactly how many people choose what age of retirement. The outliers are also quite hard to visualize in this graph.