

ADS project group6

2141

2024-03-14

Part 1

In 2019, what region of the world has the highest rate of alcohol-related deaths among men aged 40-44?

```
library(ggplot2)
library(dplyr)
# Import the data
substance_use <- read.csv("D:/R download myself/substance_use.csv")

# Starting a pipeline operation allows us to pass the output of one function sequentially to the next f
highest_alcohol_deaths <- substance_use %>%
  # Filter the data
  filter(measure == "Deaths", year == 2019, age == "40 to 44", sex == "Male", cause == "Alcohol use dis
  # Group by 'location' so that the average mortality rate is calculated separately for each region
  group_by(location) %>%
  # The data for each region were aggregated and the average mortality rate was calculated, ignoring NA
  summarize(average_death_rate = mean(val, na.rm = TRUE)) %>%
  # Rank the calculated average death rates in descending order
  arrange(desc(average_death_rate)) %>%
  # Select the region with the highest average death rate
  top_n(1, average_death_rate)
highest_alcohol_deaths
```

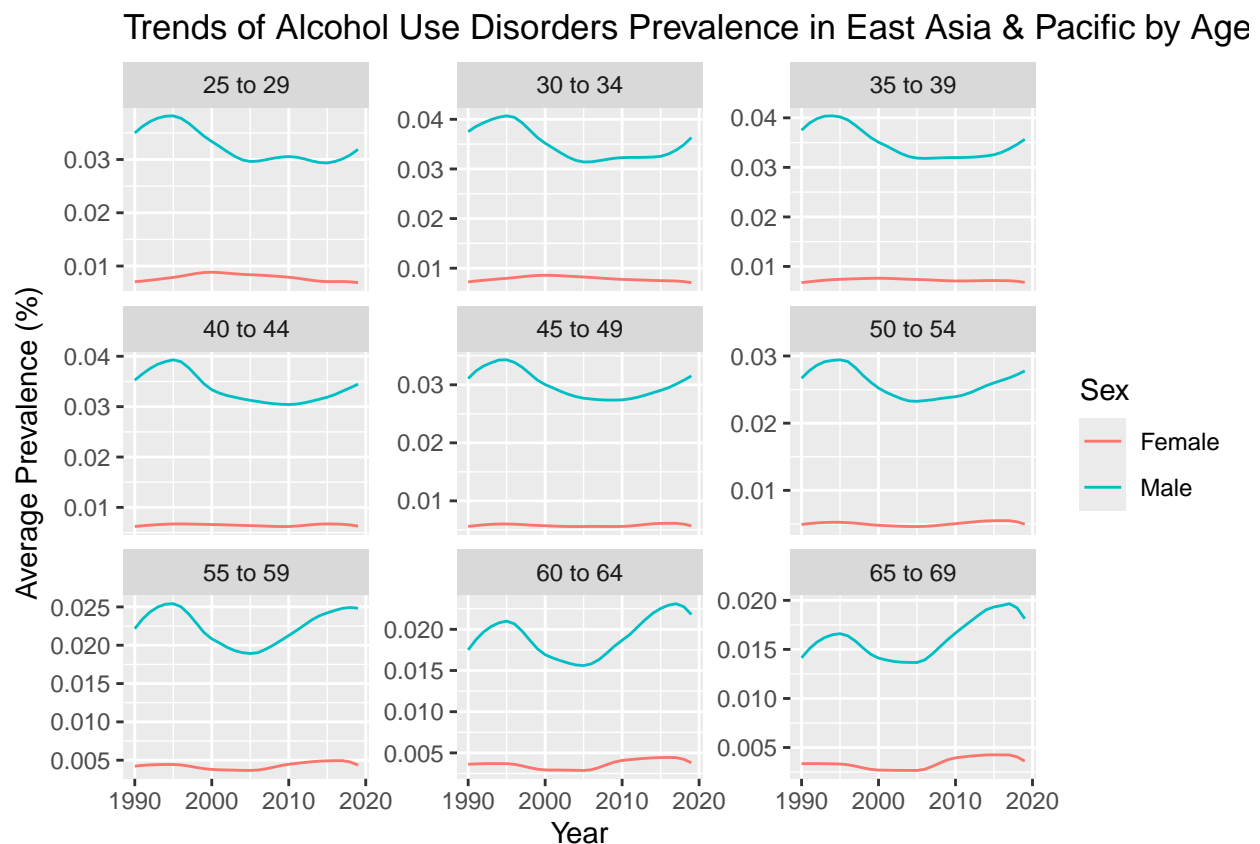
```
## # A tibble: 1 x 2
##   location                average_death_rate
##   <chr>                    <dbl>
## 1 Europe & Central Asia - WB      0.0538
```

In 2019, Europe & Central Asia has the highest rate of alcohol-related deaths among men aged 40-44.

Looking at the prevalence of alcohol-related disease in the East Asia and Pacific region, how has this changed over time and in the different age groups? Is there a difference between men and women?

```
# Filter the data
eap_alcohol_data <- substance_use %>% # Starting a pipeline operation
  filter(measure == 'Prevalence', cause == 'Alcohol use disorders', location == 'East Asia & Pacific - V
# The change of prevalence over time and age group was analyzed and compared by sex
eap_alcohol_trends <- eap_alcohol_data %>%
  group_by(year, age, sex) %>%
  summarize(average_prevalence = mean(val, na.rm = TRUE)) %>%
  ungroup() # Remove the group status
```

```
# Data visualization
# Create a line chart of the change in prevalence over time by gender and age group
ggplot(eap_alcohol_trends, aes(x = year, y = average_prevalence, color = sex, group = interaction(sex, age))) +
  geom_line() +
  facet_wrap(~age, scales = 'free_y') + # Use the panel diagram to show the different age groups
  labs(title = "Trends of Alcohol Use Disorders Prevalence in East Asia & Pacific by Age and Sex",
       x = "Year",
       y = "Average Prevalence (%)",
       color = "Sex") +
  theme(legend.position = "right")
```

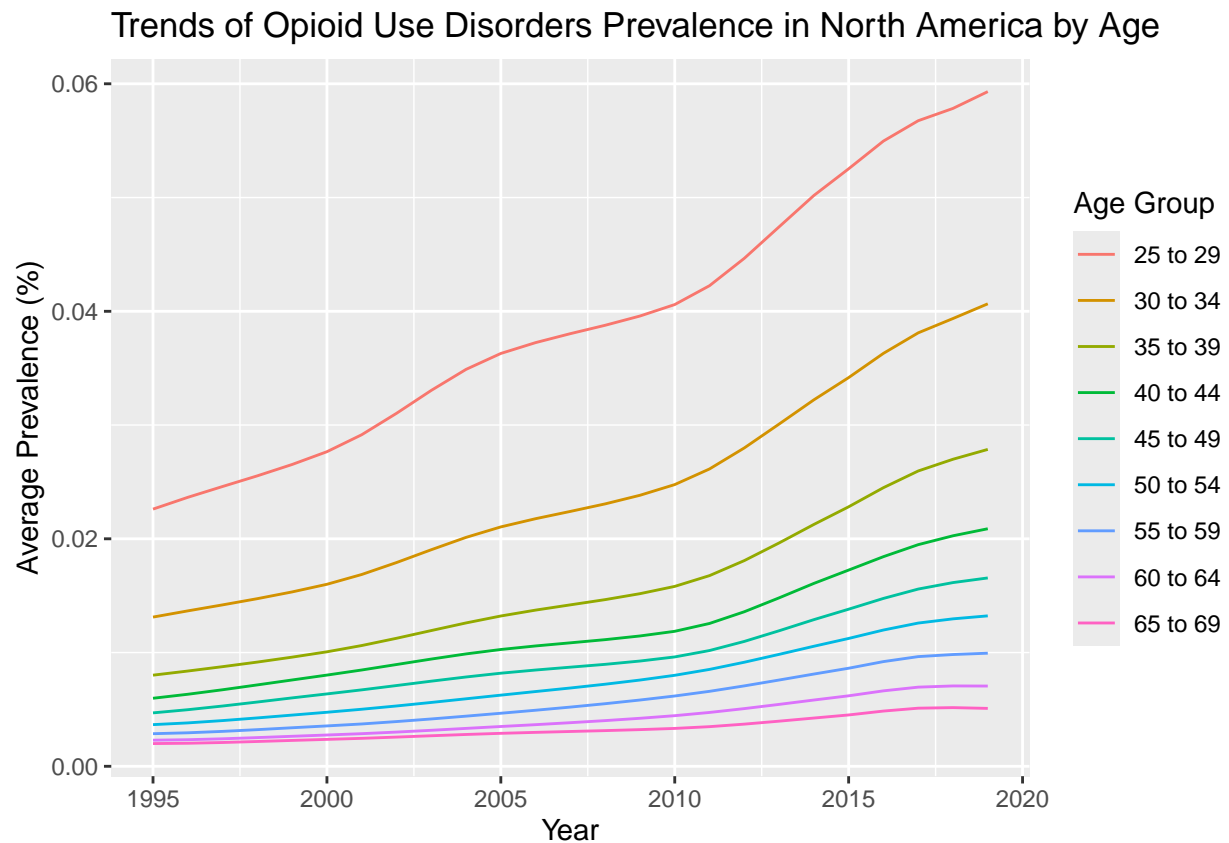


There is a significant difference between men and women.

In the United States, there is talk of an “Opioid epidemic”. Part of the problem is that since the late 1990s, doctors have increasingly been prescribing pain killers which can be highly addictive. Looking at the data from the United States, can you confirm an increase in the prevalence of diseases related to opioid use? What age group is the most affected?

```
# Filter the data
opioid_use_na <- substance_use %>% # Starting a pipeline operation
  filter(measure == 'Prevalence', location == 'North America', cause == 'Opioid use disorders', year >= 1990)
# Changes in prevalence over time and age groups were analyzed
opioid_trends_na <- opioid_use_na %>%
  group_by(year, age) %>%
  summarize(average_prevalence = mean(val, na.rm = TRUE)) %>%
  ungroup() # Remove the group status
```

```
# Data visualization
# Create a line chart that shows how the prevalence of opioid use disorder has changed over time
ggplot(opioid_trends_na, aes(x = year, y = average_prevalence, group = age)) +
  geom_line(aes(color = age)) +
  labs(title = "Trends of Opioid Use Disorders Prevalence in North America by Age",
       x = "Year",
       y = "Average Prevalence (%)",
       color = "Age Group") +
  theme(legend.position = "right")
```



There is an increase in the prevalence of diseases related to opioid use. It's the group aged 25-29 that was mostly affected