

Prison and Conviction Analysis

Kristoffer Segerstrøm, Lukas Benner and Emil Hansen

Loading and combining Excel datasets

We load multiple Excel files (from 1900 to 2024) and combine them into one dataset for analysis. We used the `file.choose()` to find out where the right excel-file was, hence the long file name.

```
file_1900 <- read_excel("/Users/lars/Desktop/Excel til eksamen /1900.xlsx")
file_1925 <- read_excel("/Users/lars/Desktop/Excel til eksamen /1925.xlsx")
file_1950 <- read_excel("/Users/lars/Desktop/Excel til eksamen /1950.xlsx")
file_1975 <- read_excel("/Users/lars/Desktop/Excel til eksamen /1975.xlsx")
file_2000 <- read_excel("/Users/lars/Desktop/Excel til eksamen /2000.xlsx")
file_2024 <- read_excel("/Users/lars/Desktop/Excel til eksamen /2024.xlsx")

# Combining them into a single data frame
combined_data <- bind_rows(file_1900, file_1925, file_1950, file_1975, file_2000, file_2024)

# Previewing the data
head(combined_data)
```

```
## # A tibble: 6 x 7
##   År Køn  Forbrydelser      Antal pr. 100.000 indbygge~1 Forbrydelse Alder
##   <dbl> <chr> <chr>          <dbl>          <dbl> <chr>      <chr>
## 1  1900 Mand Statsforbrydelser         0             0      <NA>    <NA>
## 2  1900 Mand Forbrydelser mod o~    201            8.20    <NA>    <NA>
## 3  1900 Mand Forbrydelser i emb~     4             0.163    <NA>    <NA>
## 4  1900 Mand Mened                   1             0.0408   <NA>    <NA>
## 5  1900 Mand Falsk forklaring f~    23             0.939    <NA>    <NA>
## 6  1900 Mand Forbrydelser mht. ~     0             0        <NA>    <NA>
## # i abbreviated name: 1: 'pr. 100.000 indbyggere'
```

Data preparation

We make sure that “Year” and “Convictions per 100k” are numeric and remove rows with missing values via these codes:

```
combined_data <- combined_data %>%
  mutate(
    År = as.numeric(År),
    pr_100k = as.numeric(`pr. 100.000 indbyggere`)
  ) %>%
  filter(!is.na(pr_100k))
```

Convictions over time

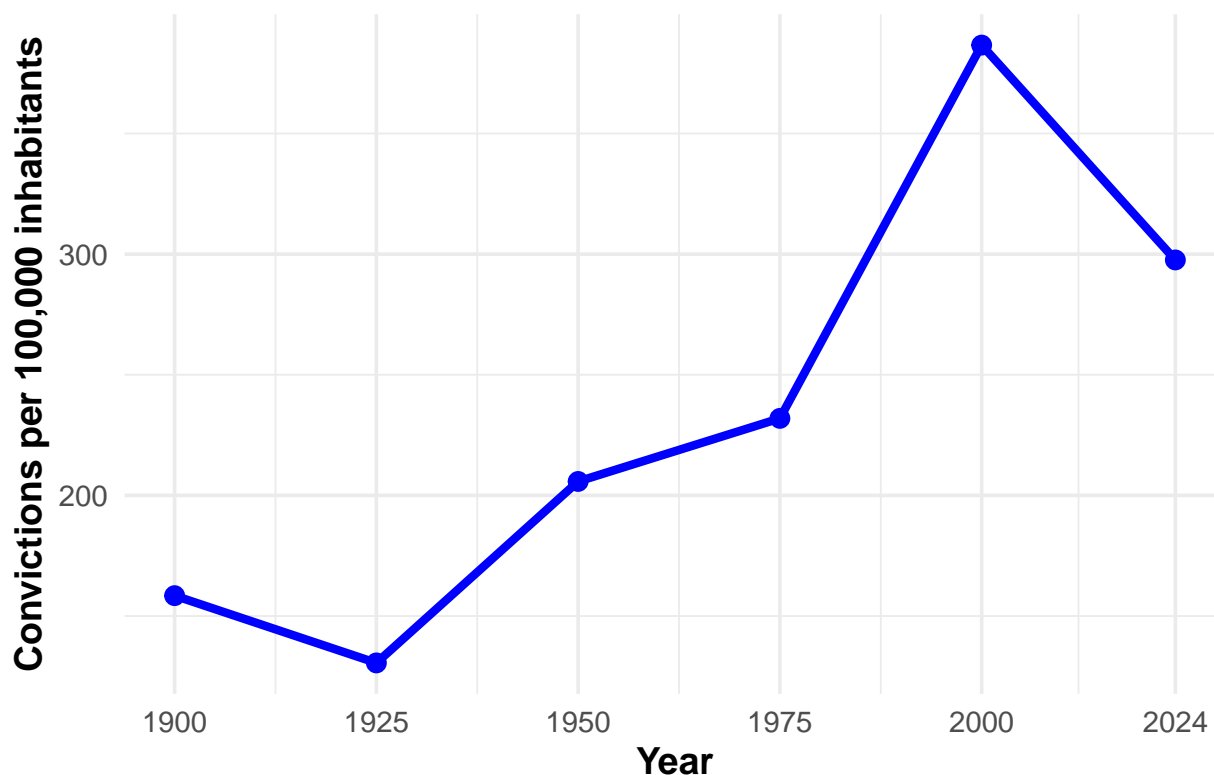
We summarize total convictions per 100,000 inhabitants per year.

```
yearly_data <- combined_data %>%
  group_by(År) %>%
  summarise(total_convicts_per_100k = sum(pr_100k, na.rm = TRUE))

# We plot the data via ggplot
# We have used the plot.title and axis.title to make our graphs and layout nicer to look at
# then we used the scale_x_continuous() to make sure we have the right data in the x-axes of the chart
ggplot(yearly_data, aes(x = År, y = total_convicts_per_100k)) +
  geom_line(color = "blue", size = 1.5) + geom_point(color = "blue", size = 3) +
  labs(
    title = "Convicted persons over time",
    x = "Year",
    y = "Convictions per 100,000 inhabitants"
  ) +
  scale_x_continuous(breaks = c(1900, 1925, 1950, 1975, 2000, 2024)) + theme_minimal(base_size = 14) +
  theme(
    plot.title = element_text(face = "bold", size = 16, hjust = 0.5),
    axis.title = element_text(face = "bold")
  )
)
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

Convicted persons over time



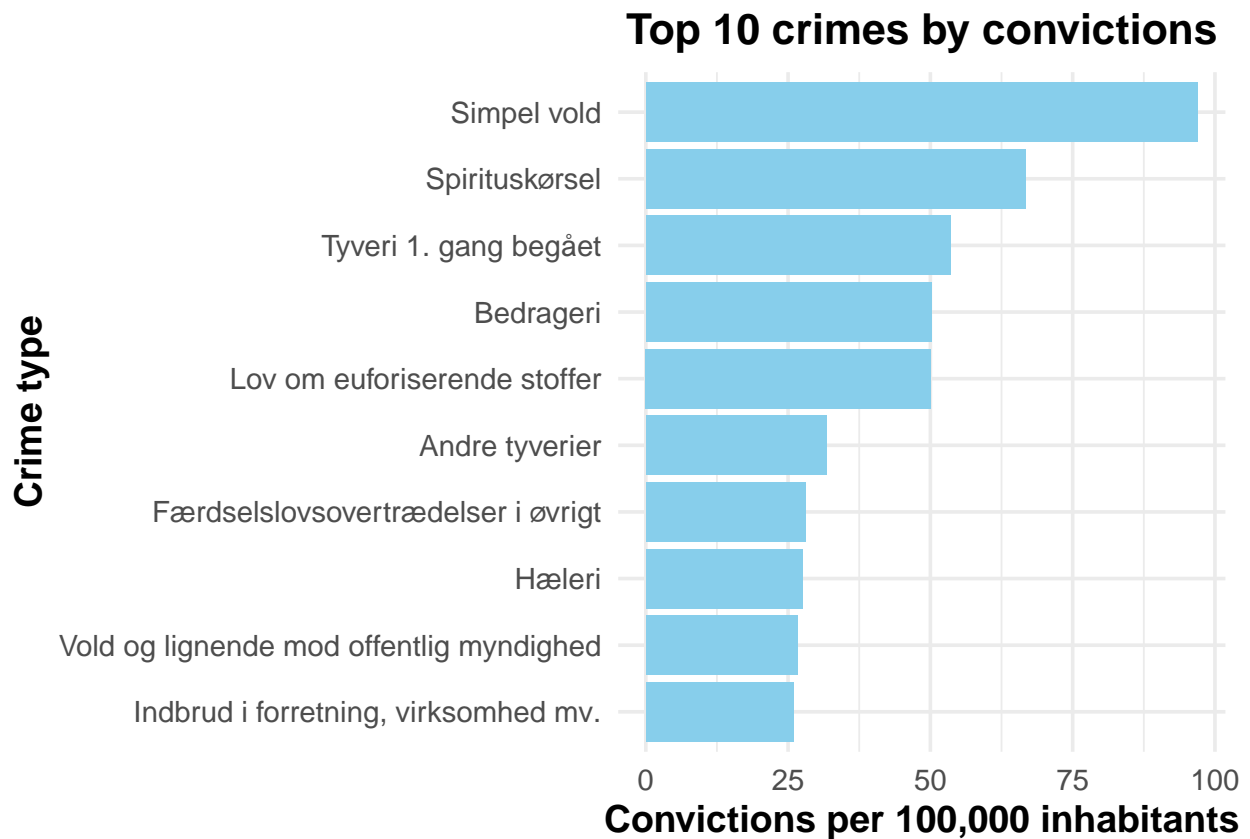
Top 10 crime categories

We identify and visualize the 10 crime types with the highest conviction rates.

```
# We will now group the combined data by crime type ("Forbrydelser"),
# then we will calculate the total conviction rate per 100,000 inhabitants for each crime
crime_data <- combined_data %>%
  group_by(Forbrydelser) %>%
  summarise(total_per_100k = sum(pr_100k, na.rm = TRUE)) %>%
  filter(!is.na(Forbrydelser), Forbrydelser != "NA") %>%
  slice_max(total_per_100k, n = 10)
#The slice_max() is used to make the top 10

# With this ggplot we are creating a horizontal bar chart showing the top 10 crime types by conviction
ggplot(crime_data, aes(x = reorder(Forbrydelser, total_per_100k), y = total_per_100k)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  coord_flip() +
  labs(
    title = "Top 10 crimes by convictions",
    x = "Crime type",
    y = "Convictions per 100,000 inhabitants"
  ) +
  theme_minimal(base_size = 14) +
  theme(
    plot.title = element_text(face = "bold", size = 16, hjust = 0.5),
```

```
axis.title = element_text(face = "bold")
)
```



Convicted by age

```
# We load the data
age_data <- read_excel("/Users/lars/Downloads/Domfældte alder.xlsx")

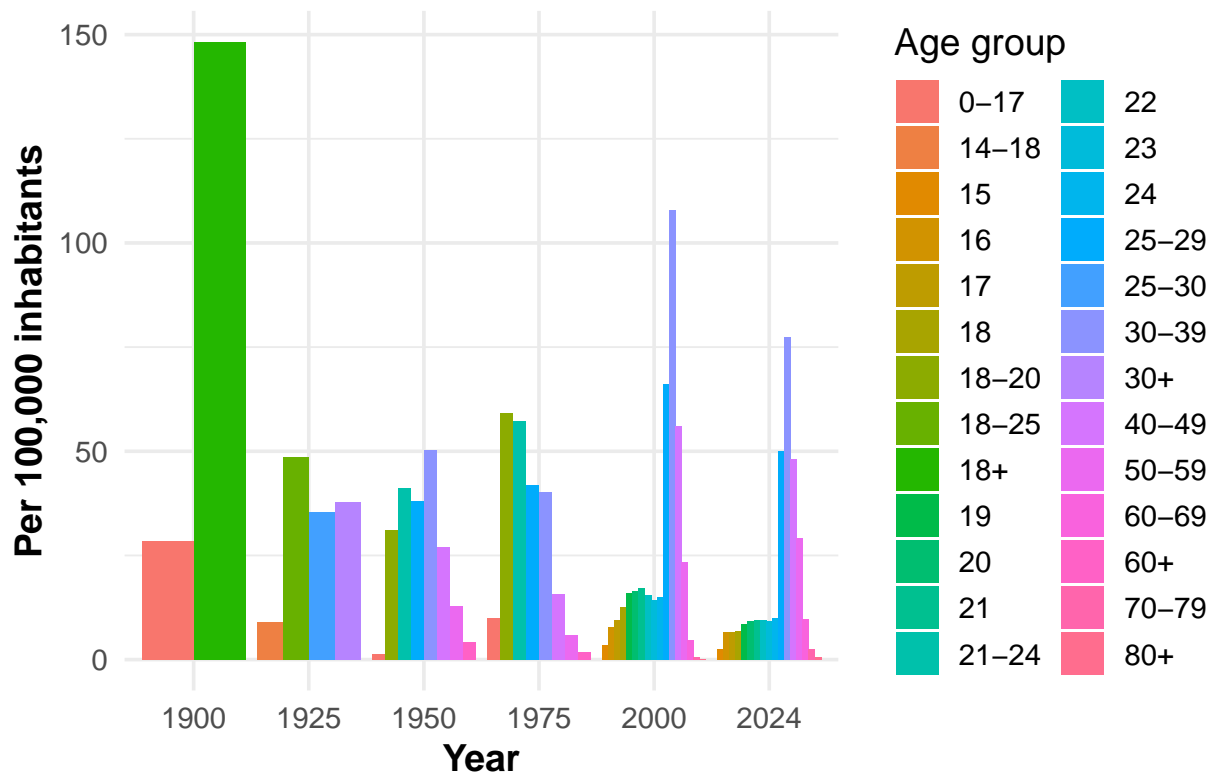
# We use the trimws() commando to clean the column names
names(age_data) <- trimws(names(age_data))

# We rename columns for consistency
age_data <- age_data %>%
  rename(
    year = Year,
    age_group = Age,
    convicted = Convicted,
    rate_per_100k = `Per 100,000 inhabitants`
  )

# We filter the data to only relevant years
age_data <- age_data %>%
  filter(year %in% c(1900, 1925, 1950, 1975, 2000, 2024))
```

```
# then we use the ggplot to make the image
ggplot(age_data, aes(x = factor(year), y = rate_per_100k, fill = age_group)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(
    title = "Convicted persons per 100,000 inhabitants by age group",
    x = "Year",
    y = "Per 100,000 inhabitants",
    fill = "Age group"
  ) +
  theme_minimal(base_size = 14) +
  theme(
    plot.title = element_text(face = "bold", size = 16, hjust = 0.5),
    axis.title = element_text(face = "bold")
  )
)
```

victed persons per 100,000 inhabitants by age group



Number of incarcerated people

```
# We load the data
incarceration_data <- read_excel("/Users/lars/Desktop/Excel til eksamen /Fængslinger i alt.xlsx")

# We use the trimws() again to once again to clean and remove extra invisible characters (whitespace) f
names(incarceration_data) <- trimws(names(incarceration_data))
```

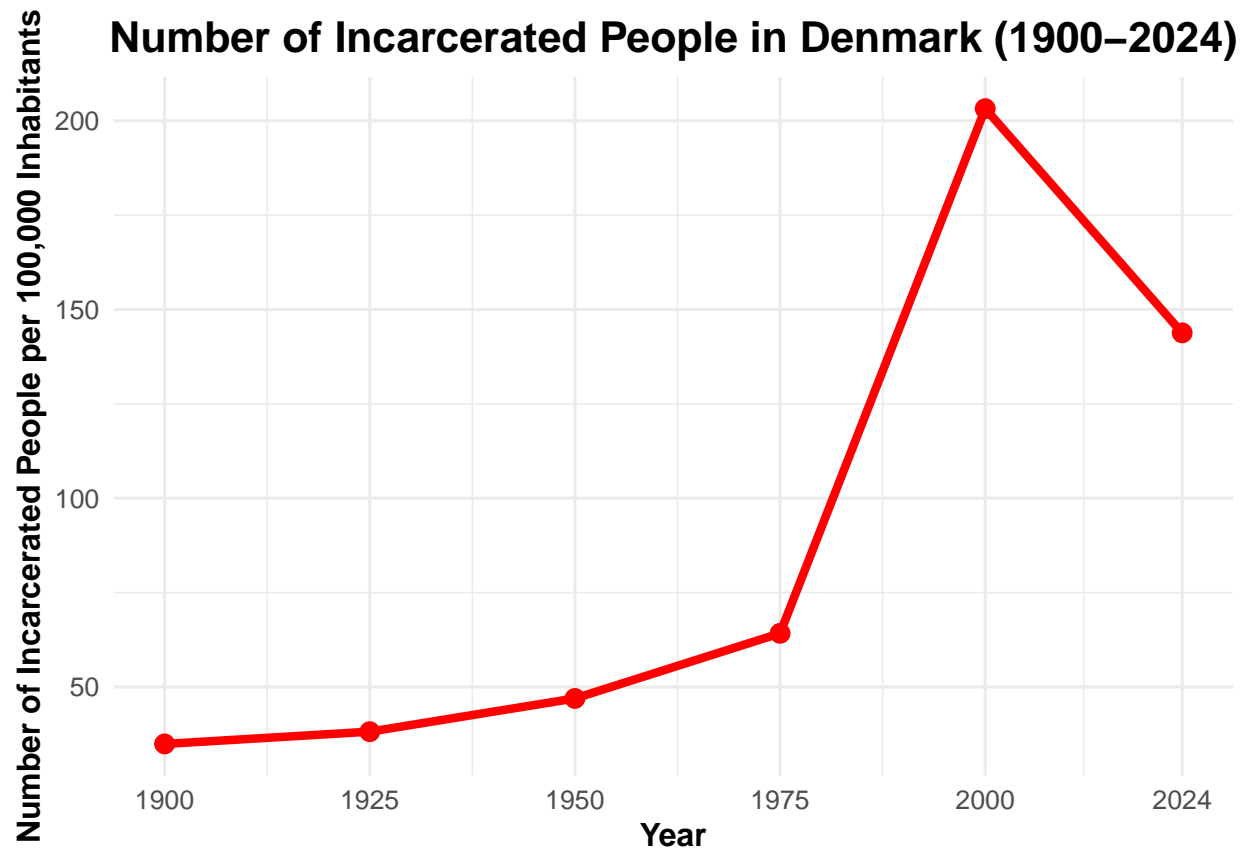
```

# We rename the columns to shorter, consistent, and English-friendly names:
incarceration_data <- incarceration_data %>%
  rename(
    year = Year,
    gender = Gender,
    count = Count,
    rate_per_100k = `Per 100,000 inhabitants`
  )

# then we summarise total incarceration rate (both genders) per year
summary_data <- incarceration_data %>%
  group_by(year) %>%
  summarise(
    total_rate_per_100k = sum(rate_per_100k)
  )

# then we create the plot
ggplot(summary_data, aes(x = year, y = total_rate_per_100k)) +
  geom_line(color = "red", linewidth = 1.5) +
  geom_point(color = "red", size = 3) + scale_x_continuous(breaks = c(1900, 1925, 1950, 1975, 2000, 2024))
labs(
  title = "Number of Incarcerated People in Denmark (1900-2024)",
  x = "Year",
  y = "Number of Incarcerated People per 100,000 Inhabitants"
) +
theme_minimal(base_size = 12) +
theme(
  plot.title = element_text(face = "bold", size = 16, hjust = 0.5),
  axis.title = element_text(face = "bold")
)

```



Number of incarcerated people divided by gender

```
# We read the excel data
incarceration_data <- read_excel("/Users/lars/Desktop/Excel til eksamen /Fængslinger i alt.xlsx")

# then we clean the column names
names(incarceration_data) <- trimws(names(incarceration_data))

# We rename the data for consistency
incarceration_data <- incarceration_data %>%
  rename(
    year = Year,
    gender = Gender,
    count = Count,
    rate_per_100k = `Per 100,000 inhabitants`
  )

# then we summarise the average rate per gender per year
summary_data <- incarceration_data %>%
  group_by(year, gender) %>%
  summarise(rate = sum(rate_per_100k), .groups = "drop")

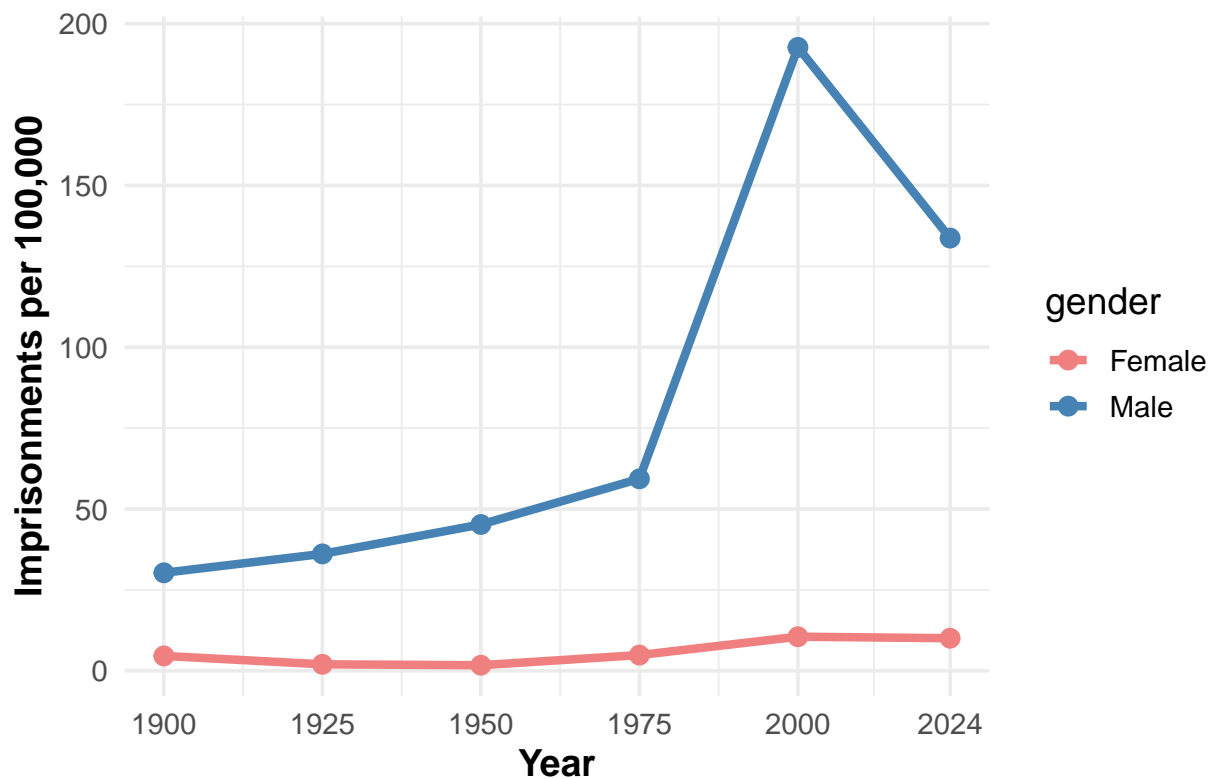
# At last we create the plot
ggplot(summary_data, aes(x = year, y = rate, color = gender)) +
```

```

geom_line(linewidth = 1.5) +
geom_point(size = 3) +
scale_x_continuous(breaks = c(1900, 1925, 1950, 1975, 2000, 2024)) +
labs(
  title = "Imprisonments per 100,000 Inhabitants Over Time by Gender",
  x = "Year",
  y = "Imprisonments per 100,000"
) +
theme_minimal(base_size = 14) +
theme(
  plot.title = element_text(face = "bold", size = 16, hjust = 0.5),
  axis.title = element_text(face = "bold")
) +
scale_color_manual(
  values = c("Male" = "steelblue", "Female" = "lightcoral")
)

```

prisonments per 100,000 Inhabitants Over Time by Gender



Animation of the gender plot

As a bonus, we have made the gender chart as a video.

```

# We start of by loading and cleaning the data
incarceration_data <- read_excel("Fængslinger i alt.xlsx", sheet = "Data-dam")

```



```

names(incarceration_data) <- trimws(names(incarceration_data))

incarceration_data <- incarceration_data %>%
  rename(
    year = Year,
    gender = Gender,
    count = Count,
    rate_per_100k = `Per 100,000 inhabitants`
  )

# then we summarise rate by gender and year
gender_summary <- incarceration_data %>%
  group_by(year, gender) %>%
  summarise(rate = sum(rate_per_100k), .groups = "drop") %>%
  filter(year %in% c(1900, 1925, 1950, 1975, 2000, 2024))

# Now we use transition_reveal to create an animated plot
p <- ggplot(gender_summary, aes(x = year, y = rate, color = gender, group = gender)) +
  geom_line(size = 1.5) +
  geom_point(size = 3) +
  scale_color_manual(values = c("Male" = "steelblue", "Female" = "lightcoral")) +
  scale_x_continuous(breaks = c(1900, 1925, 1950, 1975, 2000, 2024), limits = c(1900, 2024)) +
  labs(
    title = 'Imprisonments per 100,000 by Gender (1900-2024)',
    x = 'Year',
    y = 'Imprisonments per 100,000',
    color = 'Gender'
  ) +
  theme_minimal(base_size = 14) +
  theme(
    plot.title = element_text(face = "bold", size = 16, hjust = 0.5),
    axis.title = element_text(face = "bold")
  ) +
  transition_reveal(year)

# then we use these commands to make the animation
animate(p, fps = 10, width = 800, height = 500, renderer = av_renderer("imprisonments_by_gender.mp4"))

```

The animation created in this box will be in the “Bilag” section and in github. We were forced to write `eval = FALSE` to be able to save the file as a HTML.