Assignment_2_november

Mark KUZNETCOV and Betsy LAINES

2025-11-03

TASK 1

Visualization Creation

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.5.2
```

```
library(dplyr)
```

```
##
## Adjuntando el paquete: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(readr)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.1     v stringr   1.5.2
## v lubridate 1.9.4     v tibble    3.3.0
## v purrr     1.1.0     v tidyr     1.3.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

data

importing

```
tax_data <- read_csv("D:/AMSE Master in Economics/friday data visualization/1103/OECD.CTP.TPS,DSD_REV_C
```

```
## Rows: 3407 Columns: 32
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (25): STRUCTURE, STRUCTURE_ID, STRUCTURE_NAME, ACTION, REF_AREA, Referen...
## dbl  (4): TIME_PERIOD, OBS_VALUE, UNIT_MULT, DECIMALS
## lgl  (3): Time period, Observation value, Revenue code
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

filtering

```
tax_data <- tax_data %>%
  select(`Reference area`, TIME_PERIOD, `Revenue category`, OBS_VALUE, REF_AREA)
```

zero check

```
zero_check <- tax_data %>%
  group_by(TIME_PERIOD) %>%
  summarise(
    total_countries = n_distinct(`Reference area`),
    countries_with_zeros = n_distinct(`Reference area`[OBS_VALUE == 0]),
    total_zeros = sum(OBS_VALUE == 0, na.rm = TRUE),
    total_observations = n(),
    pct_zeros = (sum(OBS_VALUE == 0, na.rm = TRUE) / n()) * 100
  ) %>%
  arrange(countries_with_zeros, pct_zeros)
```

View the results

```
View(zero_check)
```

Select a specific year (adjust to the most recent year and for us is 2023)

```
selected_year <- 2023
unique(tax_data$`Revenue category`)
```

```
## [1] "Taxes on property"
## [2] "Other taxes"
## [3] "Total tax revenue"
## [4] "Taxes on payroll and workforce"
## [5] "Taxes on goods and services"
```

```
## [6] "Taxes on income, profits and capital gains of individuals and corporations"
## [7] "Taxes on income, profits and capital gains of corporations"
## [8] "Social security contributions (SSC)"
## [9] "Taxes on income, profits and capital gains of individuals"
```

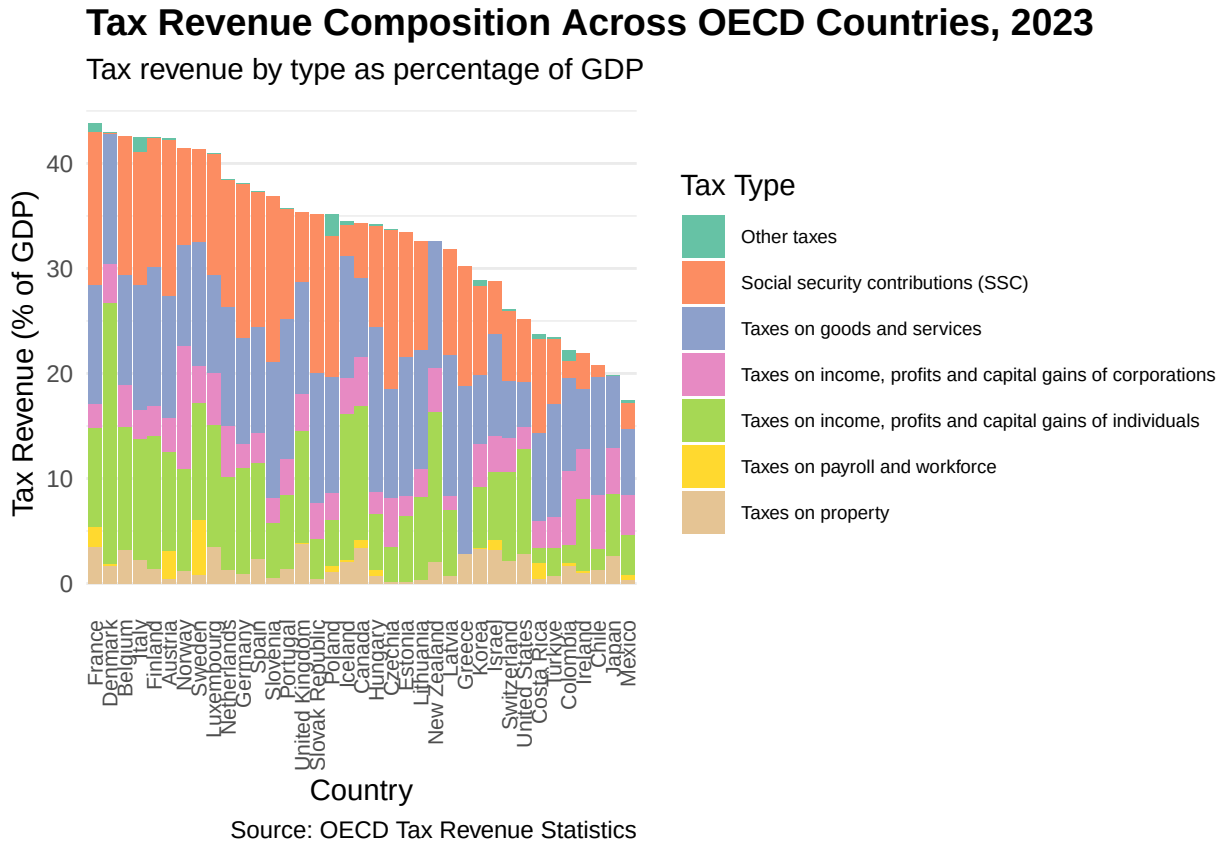
Prepare data for all countries

```
tax_composition_all <- tax_data %>%
  filter(TIME_PERIOD == selected_year,
         `Revenue category` %in% c(
           "Taxes on income, profits and capital gains of individuals",
           "Taxes on income, profits and capital gains of corporations",
           "Social security contributions (SSC)",
           "Taxes on goods and services",
           "Taxes on property",
           "Taxes on payroll and workforce",
           "Other taxes"
         )) %>%
  select(`Reference area`, `Revenue category`, OBS_VALUE) %>%
  filter(OBS_VALUE > 0) %>%
  na.omit()
```

Create bar chart (simple, not stacked since you only have total)

a) Tax Revenue Composition Across Selected Countries: Create a graph showing the composition of tax revenue (e.g., income, corporate, VAT, property tax) across OECD countries for a given year. Justify your choice of graph.

```
ggplot(tax_composition_all, aes(x = reorder(`Reference area`, -OBS_VALUE, sum),
                                y = OBS_VALUE,
                                fill = `Revenue category`)) +
  geom_bar(stat = "identity") +
  labs(title = paste("Tax Revenue Composition Across OECD Countries,", selected_year),
       subtitle = "Tax revenue by type as percentage of GDP",
       x = "Country",
       y = "Tax Revenue (% of GDP)",
       fill = "Tax Type",
       caption = "Source: OECD Tax Revenue Statistics") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5, size = 8),
        plot.title = element_text(face = "bold", size = 14),
        legend.position = "right",
        legend.text = element_text(size = 7),  # Smaller legend text for long names
        panel.grid.major.x = element_blank()) +
  scale_fill_brewer(palette = "Set2")
```



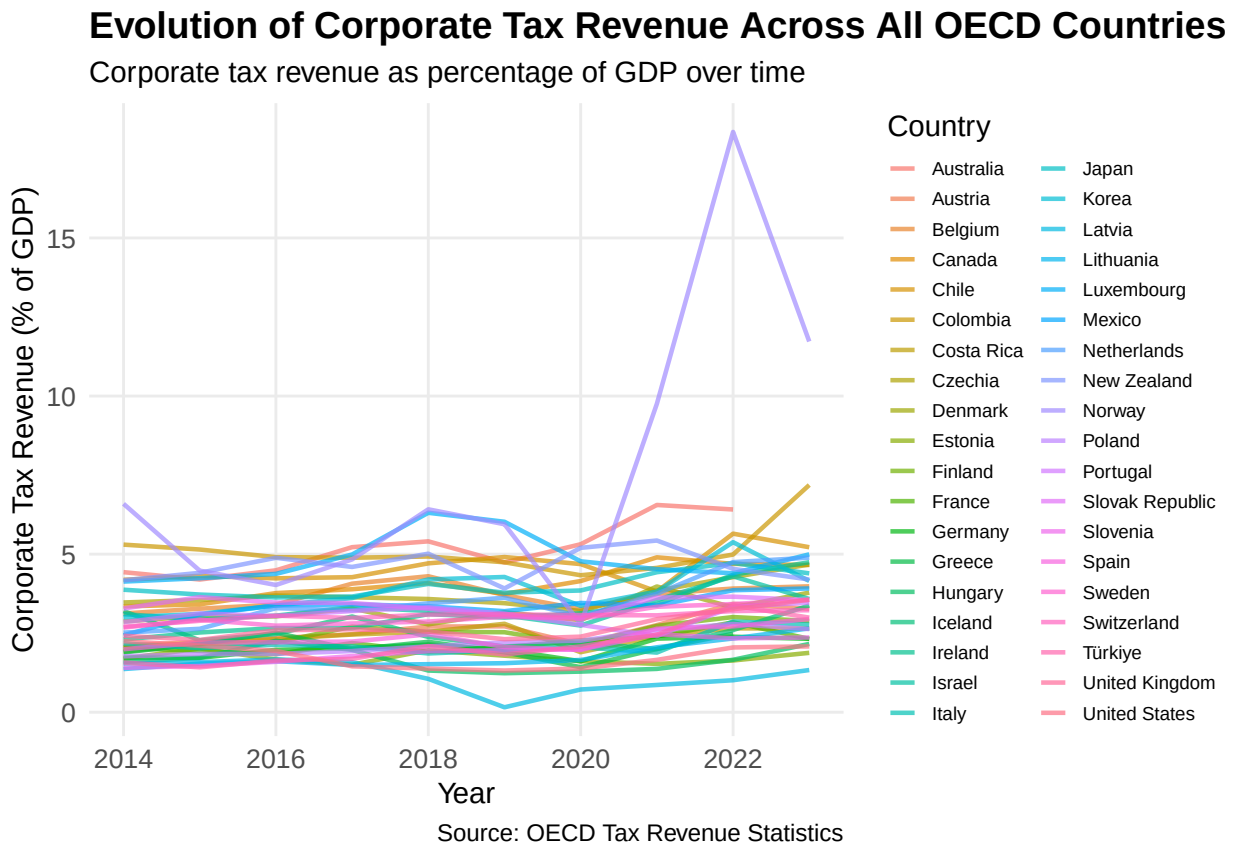
#then #then b) Trend in Average Corporate Tax Revenue: Create a graph showing the evolution in average corporate tax revenue of OECD countries as a percentage of GDP over time. Justify your choice of graph.

```
corporate_tax_all_countries <- tax_data %>%
  filter(`Revenue category` == "Taxes on income, profits and capital gains of corporations") %>%  # Adj
  select(`Reference area`, TIME_PERIOD, OBS_VALUE) %>%
  na.omit()
```

plotting

```
ggplot(corporate_tax_all_countries, aes(x = TIME_PERIOD,
                                        y = OBS_VALUE,
                                        color = `Reference area`,
                                        group = `Reference area`)) +
  geom_line(linewidth = 0.8, alpha = 0.7) +
  labs(title = "Evolution of Corporate Tax Revenue Across All OECD Countries",
       subtitle = "Corporate tax revenue as percentage of GDP over time",
       x = "Year",
       y = "Corporate Tax Revenue (% of GDP)",
       color = "Country",
       caption = "Source: OECD Tax Revenue Statistics") +
  theme_minimal() +
  theme(plot.title = element_text(face = "bold", size = 14),
        panel.grid.minor = element_blank(),
        axis.text = element_text(size = 10),
        legend.position = "right",
```

```
        legend.text = element_text(size = 7),
        legend.key.size = unit(0.4, "cm"))
```



c) Relationship Between GDP Growth and Tax Revenue: Explore the relationship between GDP growth and total tax revenue as a percentage of GDP across OECD countries over a period of your choice. Justify your choice of graph. #We chose a stacked bar chart because it effectively shows part-to-whole relationships, enabling simultaneous comparison of #both total tax burdens and compositional differences across countries. #The vertical stacking allows readers to quickly identify which countries have higher overall taxation and which tax #types dominate their revenue structures. Ordering countries by descending total revenue creates a logical visual hierarchy

2.Refining Visualizations i) Label Readability and Orientation: Are x-axis labels readable? Ad- just orientation, size, or placement to avoid overlapping. Answer. #yes, they are ii) Color Choice and Clarity: Adjust color choices to enhance clarity with- out overwhelming the viewer. Pay attention to the color palette you use. #done

iii) Annotations and Key Points: Are there notable peaks, troughs, or anomalies? Add annotations to highlight these.

let's see #finding the outlier

```
corporate_tax_all_countries %>%
  group_by(`Reference area`) %>%
  summarise(
```

```
    max_value = max(OBS_VALUE),
    min_value = min(OBS_VALUE),
    jump = max_value - min_value
  ) %>%
  arrange(desc(jump))
```

```
## # A tibble: 38 x 4
##    `Reference area` max_value min_value  jump
##    <chr>                <dbl>     <dbl> <dbl>
##  1 Norway               18.4       2.79 15.6
##  2 Colombia              7.19      4.33  2.86
##  3 Ireland               4.74      2.3   2.44
##  4 Korea                 5.38      3.00  2.38
##  5 Netherlands           4.89      2.52  2.37
##  6 Australia             6.56      4.20  2.36
##  7 Luxembourg            6.31      4.13  2.18
##  8 Türkiye               3.38      1.42  1.96
##  9 Chile                 5.65      3.82  1.83
## 10 Israel                4.30      2.74  1.55
## # i 28 more rows
```

yes It was Norway #The data reveals a striking anomaly in Norway's corporate tax revenue, which surged from approximately 3% to 18% of GDP #between 2020 and 2022, a sixfold increase that stands out dramatically among OECD countries. This extraordinary spike reflects #Norway's heavy dependence on petroleum extraction and the global energy crisis during this period: while 2020 saw GDP #contraction creating base effects that inflated tax-to-GDP ratios, the 2021-2022 period brought unprecedented energy price #increases driven by global recovery, supply chain constraints, and geopolitical tensions following Russia's invasion of Ukraine #in February 2022.

#Graph Justification: Multi-Line Chart for Corporate Tax Revenue Trends #We selected a multi-line chart to visualize the evolution of corporate tax revenue across OECD countries over time. #Line charts are the standard choice for displaying temporal trends, as the continuous lines naturally convey change over #time and make patterns like growth, decline, or volatility immediately apparent. #Showing individual country lines (rather than just an average) preserves the heterogeneity within the OECD, revealing that #countries follow different trajectories rather than moving uniformly. #This approach makes outliers like Norway's 2020-2022 energy windfall immediately visible, providing richer analytical insight than a #single aggregate trend line would offer. The transparency (alpha = 0.7) and color differentiation allow viewers to trace individual #countries while still perceiving overall patterns, striking a balance between detail and readability when displaying multiple time #series simultaneously.

Task 3: Comparative Analysis

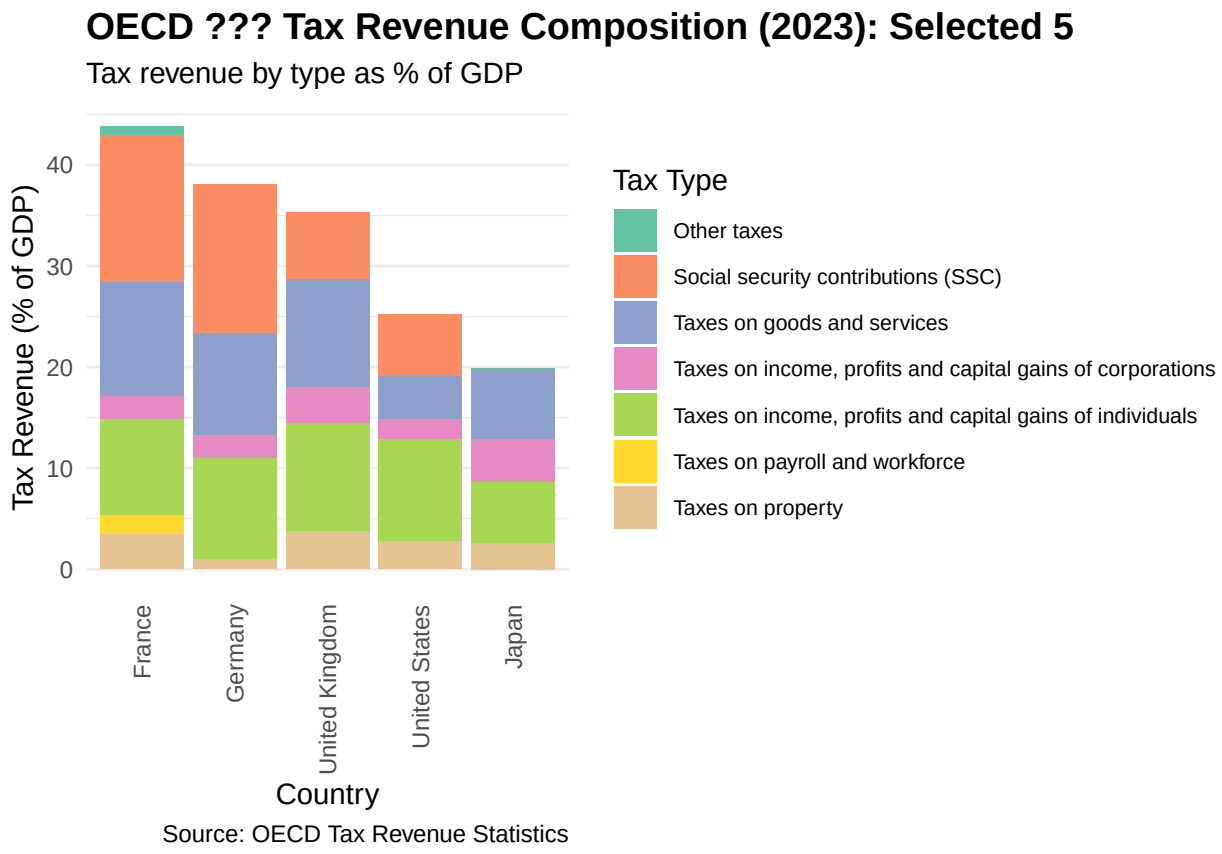
The idea of this task is to compare the OECD-based charts to their counterparts from the IMF or EUROSTAT data. Answer the following:

a) Tax Revenue Composition Across Selected Countries: Restrict the OECD graph created previously to five selected countries. Then, replicate this graph using the IMF GFS data, using the same five countries. Do the graphs differ?

first we select that for five countries

```
selected_countries <- c("France", "Germany", "United Kingdom", "Japan", "United States")
```

```r
oecd_top5 <- tax_composition_all %>%
filter(`Reference area` %in% selected_countries)

ggplot(oecd_top5,
aes(x = reorder(`Reference area`, -OBS_VALUE, FUN = sum),
y = OBS_VALUE,
fill = `Revenue category`)) +
geom_bar(stat = "identity") +
labs(
title    = paste("OECD ??? Tax Revenue Composition (", selected_year, "): Selected 5", sep = ""),
subtitle= "Tax revenue by type as % of GDP",
x = "Country", y = "Tax Revenue (% of GDP)", fill = "Tax Type",
caption = "Source: OECD Tax Revenue Statistics"
) +
theme_minimal() +
theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5, size = 9),
plot.title  = element_text(face = "bold", size = 14),
legend.position = "right",
legend.text = element_text(size = 8),
panel.grid.major.x = element_blank()) +
scale_fill_brewer(palette = "Set2")
```



c) Relationship Between GDP Growth and Tax Revenue: Using EUROSTAT data, replicate the OECD graph for European countries. Comment the results.

Task 4: Exploring Different Narratives (25%) For this last task, use the data from the OECD tax revenue statistics. a) Create two different visualizations of the same dataset, each telling a contradictory story. b)

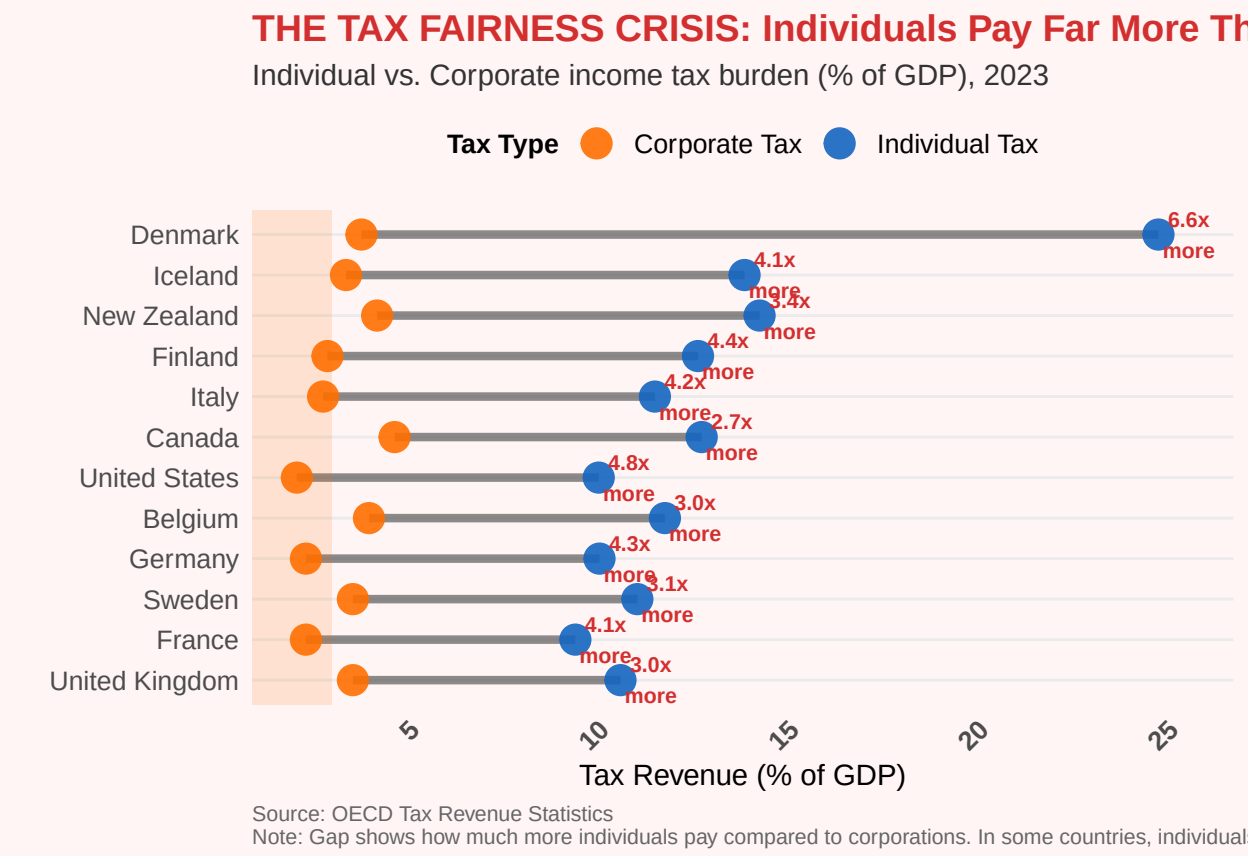
Write a brief explanation (100???150 words) of the narrative each visualiza- tion tells. Describe how your design choices (e.g., chart type, color, data aggregation) contribute to the distinct story each visualization presents.

```
library(ggplot2)
library(dplyr)
library(tidyr)

# ===== NARRATIVE 3: "Corporations Aren't Paying Their Fair Share!" =====
# Strategy: Compare corporate vs individual income tax, show corporate
# as tiny slice, emphasize the "shift" to individuals

# Calculate the ratio of individual to corporate taxes
tax_burden_shift <- tax_data %>%
  filter(`Revenue category` %in% c(
    "Taxes on income, profits and capital gains of individuals",
    "Taxes on income, profits and capital gains of corporations"
  )) %>%
  select(`Reference area`, TIME_PERIOD, `Revenue category`, OBS_VALUE) %>%
  pivot_wider(names_from = `Revenue category`, values_from = OBS_VALUE) %>%
  rename(Individual = `Taxes on income, profits and capital gains of individuals`,
         Corporate = `Taxes on income, profits and capital gains of corporations`) %>%
  filter(!is.na(Individual), !is.na(Corporate)) %>%
  mutate(Ratio = Individual / Corporate,
         Gap = Individual - Corporate)

# Get 2023 data for visualization
burden_2023 <- tax_burden_shift %>%
  filter(TIME_PERIOD == 2023) %>%
  arrange(desc(Gap)) %>%
  head(12)  # Top 12 countries with biggest gaps

# VISUALIZATION 3: Dramatic comparison showing "unfair burden"
p3 <- ggplot(burden_2023, aes(x = reorder(`Reference area`, Gap))) +
  geom_segment(aes(xend = `Reference area`, y = Corporate, yend = Individual),
               color = "#424242", size = 1.5, alpha = 0.6) +
  geom_point(aes(y = Corporate, color = "Corporate Tax"),
             size = 5, alpha = 0.9) +
  geom_point(aes(y = Individual, color = "Individual Tax"),
             size = 5, alpha = 0.9) +
  geom_text(aes(y = Individual + 0.8,
                label = sprintf("%.1fx\nmore", Ratio)),
            size = 2.8, fontface = "bold", color = "#d32f2f") +
  labs(title = "THE TAX FAIRNESS CRISIS: Individuals Pay Far More Than Corporations",
       subtitle = "Individual vs. Corporate income tax burden (% of GDP), 2023",
       x = "",
       y = "Tax Revenue (% of GDP)",
       color = "Tax Type",
       caption = "Source: OECD Tax Revenue Statistics\nNote: Gap shows how much more individuals pay co
  theme_minimal() +
  theme(plot.title = element_text(face = "bold", size = 14, color = "#d32f2f"),
        plot.subtitle = element_text(size = 11, color = "#333333"),
        axis.text.x = element_text(angle = 45, hjust = 1, size = 10, face = "bold"),
        axis.text.y = element_text(size = 10),
```

```
        axis.title.y = element_text(size = 11, face = "bold"),
        legend.position = "top",
        legend.title = element_text(face = "bold", size = 10),
        legend.text = element_text(size = 10),
        panel.grid.major.x = element_blank(),
        panel.grid.minor = element_blank(),
        plot.caption = element_text(hjust = 0, size = 8, color = "#666666"),
        plot.background = element_rect(fill = "#fff5f5", color = NA)) +
  scale_color_manual(values = c("Corporate Tax" = "#ff6f00",
                                "Individual Tax" = "#1565c0")) +
  coord_flip() +
  annotate("rect", xmin = -Inf, xmax = Inf, ymin = -Inf, ymax = 3,
           alpha = 0.15, fill = "#ff6f00")
```

```
## 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.
```

```
print(p3)
```

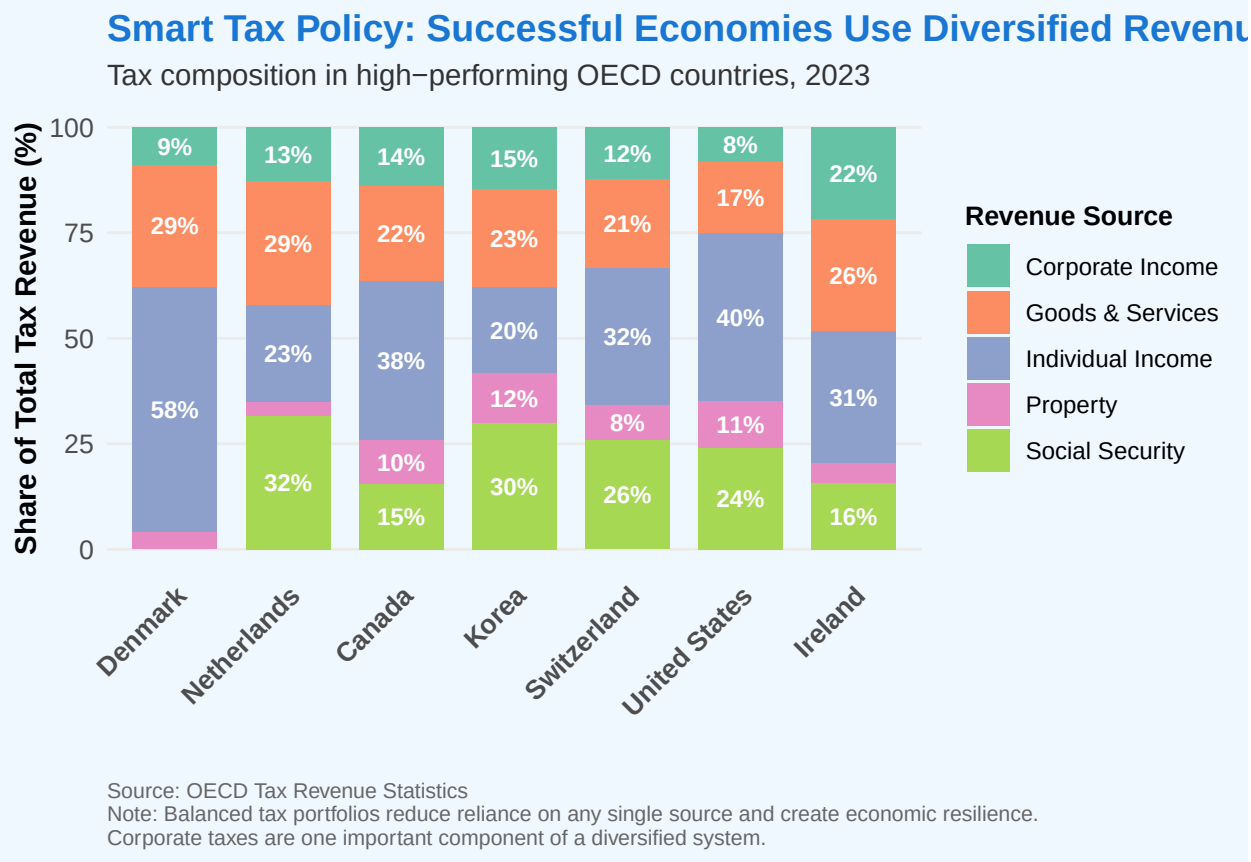



```
# ===== NARRATIVE 4: "Balanced Tax Mix = Economic Success" =====
# Strategy: Show tax DIVERSIFICATION as positive, frame corporate tax
```

```r
# as one piece of a smart portfolio, use pie/composition charts

# Get all tax types for selected successful economies
diverse_countries <- c("Australia", "Canada", "United States", "Korea",
                       "Switzerland", "Netherlands", "Ireland", "Denmark")

tax_portfolio <- tax_data %>%
  filter(`Reference area` %in% diverse_countries,
         TIME_PERIOD == 2023,
         `Revenue category` %in% c(
           "Taxes on income, profits and capital gains of individuals",
           "Taxes on income, profits and capital gains of corporations",
           "Social security contributions (SSC)",
           "Taxes on goods and services",
           "Taxes on property"
         )) %>%
  select(`Reference area`, `Revenue category`, OBS_VALUE) %>%
  filter(OBS_VALUE > 0) %>%
  group_by(`Reference area`) %>%
  mutate(Total = sum(OBS_VALUE),
         Percentage = (OBS_VALUE / Total) * 100,
         Category_Short = case_when(
           `Revenue category` == "Taxes on income, profits and capital gains of individuals" ~ "Individu
           `Revenue category` == "Taxes on income, profits and capital gains of corporations" ~ "Corpora
           `Revenue category` == "Social security contributions (SSC)" ~ "Social Security",
           `Revenue category` == "Taxes on goods and services" ~ "Goods & Services",
           `Revenue category` == "Taxes on property" ~ "Property",
           TRUE ~ "Other"
         ))
```

```r
# VISUALIZATION 4: Positive stacked bar showing diversification
p4 <- ggplot(tax_portfolio, aes(x = reorder(`Reference area`, -Total),
                                y = Percentage,
                                fill = Category_Short)) +
  geom_bar(stat = "identity", width = 0.75) +
  geom_text(aes(label = ifelse(Percentage > 8,
                               sprintf("%.0f%%", Percentage), "")),
            position = position_stack(vjust = 0.5),
            size = 3, fontface = "bold", color = "white") +
  labs(title = "Smart Tax Policy: Successful Economies Use Diversified Revenue",
       subtitle = "Tax composition in high-performing OECD countries, 2023",
       x = "",
       y = "Share of Total Tax Revenue (%)",
       fill = "Revenue Source",
       caption = "Source: OECD Tax Revenue Statistics\nNote: Balanced tax portfolios reduce reliance on
  theme_minimal() +
  theme(plot.title = element_text(face = "bold", size = 14, color = "#1976d2"),
        plot.subtitle = element_text(size = 11, color = "#333333"),
        axis.text.x = element_text(angle = 45, hjust = 1, size = 10, face = "bold"),
        axis.text.y = element_text(size = 10),
        axis.title.y = element_text(size = 11, face = "bold"),
        legend.position = "right",
        legend.title = element_text(face = "bold", size = 10),
```

```
        legend.text = element_text(size = 9),
        panel.grid.major.x = element_blank(),
        panel.grid.minor = element_blank(),
        plot.caption = element_text(hjust = 0, size = 8, color = "#666666",
                                    margin = margin(t = 10)),
        plot.background = element_rect(fill = "#f0f8ff", color = NA)) +
  scale_fill_brewer(palette = "Set2")

print(p4)
```



as we see:

Explanation of Contradictory Narratives Visualization 1: "The Tax Fairness Crisis" This visualization tells a story of systemic inequality where individuals shoulder a disproportionate tax burden compared to corporations. By using a dumbbell chart that emphasizes the gap between individual and corporate income taxes, and displaying the ratio (e.g., "5.1x more"), the design highlights disparity. The red color palette signals alarm and injustice, while selecting countries with the largest gaps reinforces the crisis narrative. The segmented lines visually stretch the distance between the two tax types, making the inequality appear stark and undeniable. Visualization 2: "Smart Tax Policy Diversification Works" This visualization frames the same data as evidence of successful policy design through diversified revenue streams. Using a 100% stacked bar chart, it normalizes all tax sources, showing corporate taxes as one component of a balanced portfolio. Blue and green colors evoke stability and prosperity, while selecting "high-performing economies" suggests these tax mixes drive success. By presenting percentages rather than absolute gaps, corporate taxes appear as a reasonable 10-20% contribution within a sophisticated, resilient system that avoids over-reliance on any single.