Final Report

Liz

2025-02-07

## Data Scraping

**Explanation**

I began by selecting the “All-Time Olympic Games Medal Table” webpage on Wikipedia as my data source. This webpage contains several tables summarizing Olympic performances of countries, including participation counts and medal tallies for the Summer, Winter, and combined Olympic Games.

To start, I saved the URL of the webpage and read its HTML content into R using the read\_html() function. This allowed me to access and work with the entire structure of the webpage. Then, I identified all the tables on the page by targeting elements with the CSS class “wikitable”, which is commonly used for structured tables on Wikipedia. Using the html\_table() function, I converted these tables into a list of data frames, with each data frame corresponding to one table on the webpage.

# Save the URL and scrape the webpage  
url <- "https://en.wikipedia.org/wiki/All-time\_Olympic\_Games\_medal\_table"  
olympic <- read\_html(x = url)  
olympic

## {html\_document}  
## <html class="client-nojs vector-feature-language-in-header-enabled vector-feature-language-in-main-page-header-disabled vector-feature-page-tools-pinned-disabled vector-feature-toc-pinned-clientpref-1 vector-feature-main-menu-pinned-disabled vector-feature-limited-width-clientpref-1 vector-feature-limited-width-content-enabled vector-feature-custom-font-size-clientpref-1 vector-feature-appearance-pinned-clientpref-1 vector-feature-night-mode-enabled skin-theme-clientpref-day vector-sticky-header-enabled vector-toc-available" lang="en" dir="ltr">  
## [1] <head>\n<meta http-equiv="Content-Type" content="text/html; charset=UTF-8 ...  
## [2] <body class="skin--responsive skin-vector skin-vector-search-vue mediawik ...

# Extract all tables with the class "wikitable"  
tables <- html\_elements(olympic, css = "table.wikitable")  
  
# Convert all table nodes into a list of data frames  
all\_tables <- html\_table(tables, fill = TRUE)

## Extracting Tables

**Explanation**

From this webpage, I identified and extracted three relevant tables for my analysis: the all-time combined Olympic medal table, the Summer Olympics medal table, and the Winter Olympics medal table.

# Extract the first table (All-time Olympic Games medal table)  
colnames(all\_tables[[1]])

## [1] "Team" "Summer Olympic Games" "Summer Olympic Games"  
## [4] "Summer Olympic Games" "Summer Olympic Games" "Summer Olympic Games"  
## [7] "Winter Olympic Games" "Winter Olympic Games" "Winter Olympic Games"  
## [10] "Winter Olympic Games" "Winter Olympic Games" "Combined total"   
## [13] "Combined total" "Combined total" "Combined total"   
## [16] "Combined total"

first\_table <- all\_tables[[1]]  
  
# Rename columns for the first table  
colnames(first\_table) <- c(  
 "Team",  
 "Summer\_No",  
 "Summer\_Gold",  
 "Summer\_Silver",  
 "Summer\_Bronze",  
 "Summer\_Total",  
 "Winter\_No",  
 "Winter\_Gold",  
 "Winter\_Silver",  
 "Winter\_Bronze",  
 "Winter\_Total",  
 "Combined\_No",  
 "Combined\_Gold",  
 "Combined\_Silver",  
 "Combined\_Bronze",  
 "Combined\_Total"  
)  
  
# Remove the first row (unwanted header)  
first\_table\_cleaned <- first\_table[-1, ] %>%  
 mutate(Team = str\_extract(Team, "^[^\\(]+\\s\*\\([^\\)]+\\)")) %>%  
 mutate(Team = str\_trim(Team))   
  
# Save the cleaned table to a CSV  
write\_csv(first\_table\_cleaned, "all\_time\_olympics\_medal\_table.csv")  
  
# Display the table  
print(first\_table\_cleaned)

## # A tibble: 163 × 16  
## Team Summer\_No Summer\_Gold Summer\_Silver Summer\_Bronze Summer\_Total  
## <chr> <chr> <chr> <chr> <chr> <chr>   
## 1 Afghanistan (… 16 0 0 2 2   
## 2 Albania (ALB) 10 0 0 2 2   
## 3 Algeria (ALG) 15 7 4 9 20   
## 4 Argentina (AR… 26 22 27 31 80   
## 5 Armenia (ARM) 8 2 11 9 22   
## 6 Australasia (… 2 3 4 5 12   
## 7 Australia (AU… 28 182 192 226 600   
## 8 Austria (AUT) 29 22 35 44 101   
## 9 Azerbaijan (A… 8 9 16 31 56   
## 10 Bahamas (BAH) 18 8 2 6 16   
## # ℹ 153 more rows  
## # ℹ 10 more variables: Winter\_No <chr>, Winter\_Gold <chr>, Winter\_Silver <chr>,  
## # Winter\_Bronze <chr>, Winter\_Total <chr>, Combined\_No <chr>,  
## # Combined\_Gold <chr>, Combined\_Silver <chr>, Combined\_Bronze <chr>,  
## # Combined\_Total <chr>

# Inspect tables to find tables I want to extract  
for (i in seq\_along(all\_tables)) {  
 print(paste("Table", i))  
 print((all\_tables[[i]]))  
}

## [1] "Table 1"  
## # A tibble: 164 × 16  
## Team `Summer Olympic Games` `Summer Olympic Games` `Summer Olympic Games`  
## <chr> <chr> <chr> <chr>   
## 1 .mw-par… No. "" ""   
## 2 Afghani… 16 "0" "0"   
## 3 Albania… 10 "0" "0"   
## 4 Algeria… 15 "7" "4"   
## 5 Argenti… 26 "22" "27"   
## 6 Armenia… 8 "2" "11"   
## 7 Austral… 2 "3" "4"   
## 8 Austral… 28 "182" "192"   
## 9 Austria… 29 "22" "35"   
## 10 Azerbai… 8 "9" "16"   
## # ℹ 154 more rows  
## # ℹ 12 more variables: `Summer Olympic Games` <chr>,  
## # `Summer Olympic Games` <chr>, `Winter Olympic Games` <chr>,  
## # `Winter Olympic Games` <chr>, `Winter Olympic Games` <chr>,  
## # `Winter Olympic Games` <chr>, `Winter Olympic Games` <chr>,  
## # `Combined total` <chr>, `Combined total` <chr>, `Combined total` <chr>,  
## # `Combined total` <chr>, `Combined total` <chr>  
## [1] "Table 2"  
## # A tibble: 71 × 4  
## `Team (IOC code)` `No. Summer` `No. Winter` `No. Games`  
## <chr> <int> <int> <int>  
## 1 American Samoa (ASA) 10 2 12  
## 2 Andorra (AND) 13 13 26  
## 3 Angola (ANG) 11 0 11  
## 4 Antigua and Barbuda (ANT) 12 0 12  
## 5 Aruba (ARU) 10 0 10  
## 6 Bangladesh (BAN) 11 0 11  
## 7 Belize (BIZ)[BIZ] 14 0 14  
## 8 Benin (BEN)[BEN] 13 0 13  
## 9 Bhutan (BHU) 11 0 11  
## 10 Bolivia (BOL) 16 7 23  
## # ℹ 61 more rows  
## [1] "Table 3"  
## # A tibble: 12 × 16  
## Team `Summer Olympic Games` `Summer Olympic Games` `Summer Olympic Games`  
## <chr> <chr> <int> <int>  
## 1 Team (I… No. NA NA  
## 2 Bohemia… 3 0 1  
## 3 British… 1 0 0  
## 4 Czechos… 16 49 49  
## 5 East Ge… 5 153 129  
## 6 West Ge… 5 56 67  
## 7 Netherl… 13 0 1  
## 8 Russian… 3 1 4  
## 9 Soviet … 9 395 319  
## 10 Serbia … 3 2 4  
## 11 Yugosla… 16 26 29  
## 12 Totals 24 682 603  
## # ℹ 12 more variables: `Summer Olympic Games` <chr>,  
## # `Summer Olympic Games` <chr>, `Winter Olympic Games` <chr>,  
## # `Winter Olympic Games` <chr>, `Winter Olympic Games` <chr>,  
## # `Winter Olympic Games` <chr>, `Winter Olympic Games` <chr>,  
## # `Combined total` <chr>, `Combined total` <chr>, `Combined total` <chr>,  
## # `Combined total` <chr>, `Combined total` <chr>  
## [1] "Table 4"  
## # A tibble: 12 × 16  
## Team `Summer Olympic Games` `Summer Olympic Games` `Summer Olympic Games`  
## <chr> <chr> <int> <int>  
## 1 Team (I… No. NA NA  
## 2 Austral… 2 3 4  
## 3 Individ… 1 1 3  
## 4 Refugee… 3 0 0  
## 5 United … 3 28 54  
## 6 Unified… 1 45 38  
## 7 Olympic… 0 0 0  
## 8 ROC (RO… 1 20 28  
## 9 Indepen… 3 1 0  
## 10 Indepen… 1 0 1  
## 11 Mixed t… 3 11 6  
## 12 Totals 18 109 134  
## # ℹ 12 more variables: `Summer Olympic Games` <chr>,  
## # `Summer Olympic Games` <chr>, `Winter Olympic Games` <chr>,  
## # `Winter Olympic Games` <chr>, `Winter Olympic Games` <chr>,  
## # `Winter Olympic Games` <chr>, `Winter Olympic Games` <chr>,  
## # `Combined total` <chr>, `Combined total` <chr>, `Combined total` <chr>,  
## # `Combined total` <chr>, `Combined total` <chr>  
## [1] "Table 5"  
## # A tibble: 10 × 6  
## No. Nation Gold Silver Bronze Total  
## <int> <chr> <chr> <int> <int> <chr>  
## 1 1 United States (USA) 1,105 879 781 2,765  
## 2 2 Russia (RUS)[I] 608 514 501 1,623  
## 3 3 Germany (GER)[II] 455 470 499 1,424  
## 4 4 China (CHN) 303 226 198 727   
## 5 5 Great Britain (GBR) 298 339 343 980   
## 6 6 France (FRA) 239 279 306 821   
## 7 7 Italy (ITA) 229 201 221 651   
## 8 8 Japan (JPN) 189 162 191 542   
## 9 9 Hungary (HUN) 187 161 182 530   
## 10 10 Australia (AUS) 182 192 226 600   
## [1] "Table 6"  
## # A tibble: 10 × 6  
## No. Nation Gold Silver Bronze Total  
## <int> <chr> <chr> <int> <int> <chr>  
## 1 1 United States (USA) 1,105 879 781 2,765  
## 2 2 Soviet Union (URS) 395 319 296 1,010  
## 3 3 China (CHN) 303 226 198 727   
## 4 4 Great Britain (GBR) 298 339 343 980   
## 5 5 France (FRA) 239 279 306 821   
## 6 6 Italy (ITA) 229 201 221 651   
## 7 7 Germany (GER) 218 220 255 693   
## 8 8 Japan (JPN) 189 162 191 542   
## 9 9 Hungary (HUN) 187 161 182 530   
## 10 10 Australia (AUS) 182 192 226 600   
## [1] "Table 7"  
## # A tibble: 10 × 6  
## No. Nation Gold Silver Bronze Total  
## <int> <chr> <int> <int> <int> <int>  
## 1 1 Germany (GER)[I] 162 155 118 435  
## 2 2 Norway (NOR) 148 134 123 405  
## 3 3 Russia (RUS)[II] 140 120 126 386  
## 4 4 United States (USA) 114 121 95 330  
## 5 5 Canada (CAN) 77 72 76 225  
## 6 6 Austria (AUT) 71 88 91 250  
## 7 7 Sweden (SWE) 65 51 60 176  
## 8 8 Switzerland (SUI) 63 47 58 168  
## 9 9 Netherlands (NED) 53 49 45 147  
## 10 10 Finland (FIN) 45 65 65 175  
## [1] "Table 8"  
## # A tibble: 10 × 6  
## No. Nation Gold Silver Bronze Total  
## <int> <chr> <int> <int> <int> <int>  
## 1 1 Norway (NOR) 148 134 123 405  
## 2 2 United States (USA) 114 121 95 330  
## 3 3 Germany (GER) 104 98 65 267  
## 4 4 Soviet Union (URS) 78 57 59 194  
## 5 5 Canada (CAN) 77 72 76 225  
## 6 6 Austria (AUT) 71 88 91 250  
## 7 7 Sweden (SWE) 65 51 60 176  
## 8 8 Switzerland (SUI) 63 47 58 168  
## 9 9 Netherlands (NED) 53 49 45 147  
## 10 10 Russia (RUS) 47 39 35 121  
## [1] "Table 9"  
## # A tibble: 10 × 6  
## No. Nation Gold Silver Bronze Total  
## <int> <chr> <chr> <chr> <int> <chr>  
## 1 1 United States (USA) 1,219 1,000 876 3,095  
## 2 2 Russia (RUS)[I] 748 634 627 2,009  
## 3 3 Germany (GER)[II] 617 625 617 1,859  
## 4 4 China (CHN) 325 258 221 804   
## 5 5 Great Britain (GBR) 310 344 360 1,014  
## 6 6 France (FRA) 280 320 354 954   
## 7 7 Italy (ITA) 271 244 284 799   
## 8 8 Sweden (SWE) 216 232 242 690   
## 9 9 Norway (NOR) 213 187 176 576   
## 10 10 Japan (JPN) 206 191 221 618   
## [1] "Table 10"  
## # A tibble: 10 × 6  
## No. Nation Gold Silver Bronze Total  
## <int> <chr> <chr> <chr> <int> <chr>  
## 1 1 United States (USA) 1,219 1,000 876 3,095  
## 2 2 Soviet Union (URS) 473 376 355 1,204  
## 3 3 China (CHN) 325 258 221 804   
## 4 4 Germany (GER) 322 318 320 960   
## 5 5 Great Britain (GBR) 310 344 360 1,014  
## 6 6 France (FRA) 281 320 360 961   
## 7 7 Italy (ITA) 271 244 284 799   
## 8 8 Sweden (SWE) 216 232 242 690   
## 9 9 Norway (NOR) 213 187 176 576   
## 10 10 Japan (JPN) 206 191 221 618   
## [1] "Table 11"  
## # A tibble: 161 × 6  
## Rank NOC Gold Silver Bronze Total  
## <chr> <chr> <chr> <chr> <chr> <chr>  
## 1 1 United States 1,105 879 781 2,765  
## 2 2 Soviet Union\* 395 319 296 1,010  
## 3 3 China 303 226 198 727   
## 4 4 Great Britain 298 339 343 980   
## 5 5 France 239 278 299 816   
## 6 6 Italy 229 201 228 658   
## 7 7 Germany 218 220 255 693   
## 8 8 Japan 189 162 191 542   
## 9 9 Hungary 187 161 182 530   
## 10 10 Australia 182 192 226 600   
## # ℹ 151 more rows  
## [1] "Table 12"  
## # A tibble: 48 × 6  
## Rank NOC Gold Silver Bronze Total  
## <chr> <chr> <chr> <chr> <chr> <chr>  
## 1 1 Norway 148 134 123 405   
## 2 2 United States 114 121 95 330   
## 3 3 Germany 104 98 65 267   
## 4 4 Soviet Union\* 78 57 59 194   
## 5 5 Canada 77 72 76 225   
## 6 6 Austria 71 88 91 250   
## 7 7 Sweden 65 51 60 176   
## 8 8 Switzerland 63 47 58 168   
## 9 9 Netherlands 53 49 45 147   
## 10 10 Russia 47 39 35 121   
## # ℹ 38 more rows  
## [1] "Table 13"  
## # A tibble: 163 × 6  
## Rank NOC Gold Silver Bronze Total  
## <chr> <chr> <chr> <chr> <chr> <chr>  
## 1 1 United States 1,219 1,000 876 3,095  
## 2 2 Soviet Union\* 473 376 355 1,204  
## 3 3 China 325 258 221 804   
## 4 4 Germany 322 318 320 960   
## 5 5 Great Britain 310 344 360 1,014  
## 6 6 France 281 320 360 961   
## 7 7 Italy 271 244 284 799   
## 8 8 Sweden 216 232 242 690   
## 9 9 Norway 213 187 176 576   
## 10 10 Japan 206 191 221 618   
## # ℹ 153 more rows  
## [1] "Table 14"  
## # A tibble: 2 × 1  
## X1   
## <chr>   
## 1 "Summer Olympics medal table leaders by year"   
## 2 ".mw-parser-output .div-col{margin-top:0.3em;column-width:30em}.mw-parser-out…  
## [1] "Table 15"  
## # A tibble: 7 × 3  
## Rank Country `Number of games`  
## <int> <chr> <chr>   
## 1 1 United States (USA) 19 times   
## 2 2 Soviet Union (URS) 6 times   
## 3 3 China (CHN) 1 time   
## 4 3 France (FRA) 1 time   
## 5 3 Great Britain (GBR) 1 time   
## 6 3 Germany (GER) 1 time   
## 7 3 Unified Team (EUN) 1 time   
## [1] "Table 16"  
## # A tibble: 2 × 1  
## X1   
## <chr>   
## 1 "Winter Olympics medal table leaders by year"   
## 2 "1924:  Norway\n1928:  Norway\n1932:  United States\n1936:  Norway\n1948:  No…  
## [1] "Table 17"  
## # A tibble: 8 × 3  
## Rank Country `Number of games`  
## <int> <chr> <chr>   
## 1 1 Norway (NOR) 10 times   
## 2 2 Soviet Union (URS) 7 times   
## 3 3 Germany (GER) 3 times   
## 4 4 United States (USA) 1 time   
## 5 4 Sweden (SWE) 1 time   
## 6 4 East Germany (GDR) 1 time   
## 7 4 Canada (CAN) 1 time   
## 8 4 Russia (RUS) 1 time   
## [1] "Table 18"  
## # A tibble: 3 × 3  
## Date Team Team   
## <chr> <chr> <chr>   
## 1 1896–1904 Australia (AUS) ""   
## 2 1908–1912 Australasia (ANZ) "Australasia (ANZ)"  
## 3 1920– Australia (AUS) "New Zealand (NZL)"  
## [1] "Table 19"  
## # A tibble: 5 × 16  
## `` `Summer Games` `Summer Games` `Summer Games` `Summer Games`  
## <chr> <chr> <int> <int> <int>  
## 1 Team (IOC code) No. NA NA NA  
## 2 Australasia (ANZ) 2 3 4 5  
## 3 Australia (AUS) 28 182 192 226  
## 4 New Zealand (NZL) 25 63 40 54  
## 5 Total 30 248 236 285  
## # ℹ 11 more variables: `Summer Games` <chr>, `Winter Games` <chr>,  
## # `Winter Games` <chr>, `Winter Games` <chr>, `Winter Games` <chr>,  
## # `Winter Games` <chr>, `Combined total` <chr>, `Combined total` <chr>,  
## # `Combined total` <chr>, `Combined total` <chr>, `Combined total` <chr>  
## [1] "Table 20"  
## # A tibble: 4 × 4  
## Date Team Team Team   
## <chr> <chr> <chr> <chr>   
## 1 1948–1956 Jamaica (JAM) Trinidad and Tobago (TTO) ""   
## 2 1960 British West Indies (BWI) British West Indies (BWI) "British West I…  
## 3 1964 Jamaica (JAM) Trinidad and Tobago (TTO) ""   
## 4 1968– Jamaica (JAM) Trinidad and Tobago (TTO) "Barbados (BAR)"  
## [1] "Table 21"  
## # A tibble: 6 × 16  
## `` `Summer Games` `Summer Games` `Summer Games` `Summer Games`  
## <chr> <chr> <int> <int> <int>  
## 1 Team (IOC code) No. NA NA NA  
## 2 British West Indi… 1 0 0 2  
## 3 Jamaica (JAM) 19 27 39 28  
## 4 Trinidad and Toba… 19 3 5 11  
## 5 Barbados (BAR) 14 0 0 1  
## 6 Total 19 30 44 42  
## # ℹ 11 more variables: `Summer Games` <chr>, `Winter Games` <chr>,  
## # `Winter Games` <chr>, `Winter Games` <chr>, `Winter Games` <chr>,  
## # `Winter Games` <chr>, `Combined total` <chr>, `Combined total` <chr>,  
## # `Combined total` <chr>, `Combined total` <chr>, `Combined total` <chr>  
## [1] "Table 22"  
## # A tibble: 4 × 3  
## Date Team Team   
## <chr> <chr> <chr>   
## 1 1896 "" as part of  Hungary   
## 2 1900–1912 "as  Bohemia (BOH)" as part of  Hungary   
## 3 1920–1992 "Czechoslovakia (TCH)" Czechoslovakia (TCH)  
## 4 1996– "Czech Republic (CZE)" Slovakia (SVK)   
## [1] "Table 23"  
## # A tibble: 6 × 16  
## `` `Summer Games` `Summer Games` `Summer Games` `Summer Games`  
## <chr> <chr> <int> <int> <int>  
## 1 Team (IOC code) No. NA NA NA  
## 2 Bohemia (BOH) 3 0 1 3  
## 3 Czechoslovakia (T… 16 49 49 45  
## 4 Czech Republic (C… 8 22 22 28  
## 5 Slovakia (SVK) 8 10 14 9  
## 6 Total 27 81 86 85  
## # ℹ 11 more variables: `Summer Games` <chr>, `Winter Games` <chr>,  
## # `Winter Games` <chr>, `Winter Games` <chr>, `Winter Games` <chr>,  
## # `Winter Games` <chr>, `Combined total` <chr>, `Combined total` <chr>,  
## # `Combined total` <chr>, `Combined total` <chr>, `Combined total` <chr>  
## [1] "Table 24"  
## # A tibble: 8 × 4  
## Date Team Team Team   
## <chr> <chr> <chr> <chr>   
## 1 1896–1912 Germany (GER) Germany (GER) Germany (…  
## 2 1920–1924 banned banned banned   
## 3 1928–1936 Germany (GER) Germany (GER) Germany (…  
## 4 1948 banned banned banned   
## 5 1952 Saar (SAA) Germany (GER) Germany (…  
## 6 1956–1964 United Team of Germany (EUA) United Team of Germany (EUA) United Te…  
## 7 1968–1988 West Germany (FRG) West Germany (FRG) East Germ…  
## 8 1992– Germany (GER) Germany (GER) Germany (…  
## [1] "Table 25"  
## # A tibble: 7 × 16  
## `` `Summer Games` `Summer Games` `Summer Games` `Summer Games`  
## <chr> <chr> <int> <int> <int>  
## 1 Team (IOC code) No. NA NA NA  
## 2 Germany (GER) 18 218 220 255  
## 3 Saar (SAA) 1 0 0 0  
## 4 United Team of Ge… 3 28 54 36  
## 5 East Germany (GDR) 5 153 129 127  
## 6 West Germany (FRG) 5 56 67 81  
## 7 Total 27 455 470 499  
## # ℹ 11 more variables: `Summer Games` <chr>, `Winter Games` <chr>,  
## # `Winter Games` <chr>, `Winter Games` <chr>, `Winter Games` <chr>,  
## # `Winter Games` <chr>, `Combined total` <chr>, `Combined total` <chr>,  
## # `Combined total` <chr>, `Combined total` <chr>, `Combined total` <chr>  
## [1] "Table 26"  
## # A tibble: 3 × 2  
## Date Team   
## <chr> <chr>   
## 1 1968–2012 Kuwait (KUW)   
## 2 2016 Independent Olympic Athletes (IOA) (2016)  
## 3 2020– Kuwait (KUW)   
## [1] "Table 27"  
## # A tibble: 4 × 16  
## `` `Summer Games` `Summer Games` `Summer Games` `Summer Games`  
## <chr> <chr> <int> <int> <int>  
## 1 Team (IOC code) No. NA NA NA  
## 2 Kuwait (KUW) 14 0 0 3  
## 3 Independent Olymp… 1 1 0 1  
## 4 Total 15 1 0 4  
## # ℹ 11 more variables: `Summer Games` <chr>, `Winter Games` <chr>,  
## # `Winter Games` <chr>, `Winter Games` <chr>, `Winter Games` <chr>,  
## # `Winter Games` <chr>, `Combined total` <chr>, `Combined total` <chr>,  
## # `Combined total` <chr>, `Combined total` <chr>, `Combined total` <chr>  
## [1] "Table 28"  
## # A tibble: 5 × 4  
## Date Team Team Team   
## <chr> <chr> <chr> <chr>  
## 1 1900–1948 Netherlands (NED) "" ""   
## 2 1952–1984 Netherlands (NED) "Netherlands Antilles (AHO)" "Net…  
## 3 1988–2008 Netherlands (NED) "Netherlands Antilles (AHO)" "Aru…  
## 4 2012 Netherlands (NED) "as part of  Netherlands / Independent Olym… "Aru…  
## 5 2014– Netherlands (NED) "Netherlands (NED)" "Aru…  
## [1] "Table 29"  
## # A tibble: 6 × 16  
## `` `Summer Games` `Summer Games` `Summer Games` `Summer Games`  
## <chr> <chr> <int> <int> <int>  
## 1 Team (IOC code) No. NA NA NA  
## 2 Netherlands (NED) 28 110 112 134  
## 3 Netherlands Antil… 13 0 1 0  
## 4 Aruba (ARU) 10 0 0 0  
## 5 Independent Olymp… 1 0 0 0  
## 6 Total 28 110 113 134  
## # ℹ 11 more variables: `Summer Games` <chr>, `Winter Games` <chr>,  
## # `Winter Games` <chr>, `Winter Games` <chr>, `Winter Games` <chr>,  
## # `Winter Games` <chr>, `Combined total` <chr>, `Combined total` <chr>,  
## # `Combined total` <chr>, `Combined total` <chr>, `Combined total` <chr>  
## [1] "Table 30"  
## # A tibble: 4 × 4  
## Date Team Team Team   
## <chr> <chr> <chr> <chr>   
## 1 1924–1948 Republic of China (ROC) "Republic of China (ROC)" ""   
## 2 1952 China (CHN) "" "Hong Kong (HKG)"  
## 3 1956–1996 China (CHN) "Chinese Taipei (TPE)" "Hong Kong (HKG)"  
## 4 2000– China (CHN) "Chinese Taipei (TPE)" "Hong Kong (HKG)"  
## [1] "Table 31"  
## # A tibble: 6 × 16  
## `` `Summer Games` `Summer Games` `Summer Games` `Summer Games`  
## <chr> <chr> <int> <int> <int>  
## 1 Team (IOC code) No. NA NA NA  
## 2 Republic of China… 3 0 0 0  
## 3 China (CHN) 12 303 226 198  
## 4 Chinese Taipei (T… 16 9 11 23  
## 5 Hong Kong (HKG) 18 4 3 6  
## 6 Total 20 316 240 227  
## # ℹ 11 more variables: `Summer Games` <chr>, `Winter Games` <chr>,  
## # `Winter Games` <chr>, `Winter Games` <chr>, `Winter Games` <chr>,  
## # `Winter Games` <chr>, `Combined total` <chr>, `Combined total` <chr>,  
## # `Combined total` <chr>, `Combined total` <chr>, `Combined total` <chr>  
## [1] "Table 32"  
## # A tibble: 10 × 8  
## Date Team Team Team Team Team Team Team   
## <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr>  
## 1 1900–1912 Russian Empire (RU1) "Russian Empire… "Rus… "Rus… "Rus… "Rus… "Rus…  
## 2 1920 Estonia (EST) "" "" "" "" "" ""   
## 3 1924–1936 Estonia (EST) "Latvia (LAT)" "Lit… "" "" "" ""   
## 4 1952–1988 Soviet Union (URS) "Soviet Union (… "Sov… "Sov… "Sov… "Sov… "Sov…  
## 5 1992 Estonia (EST) "Latvia (LAT)" "Lit… "Uni… "Uni… "Uni… "Uni…  
## 6 1994 Estonia (EST) "Latvia (LAT)" "Lit… "Rus… "Bel… "Arm… ""   
## 7 1996–2016 Estonia (EST) "Latvia (LAT)" "Lit… "Rus… "Bel… "Arm… "Aze…  
## 8 2018 Estonia (EST) "Latvia (LAT)" "Lit… "Oly… "Bel… "Arm… "Aze…  
## 9 2020–2022 Estonia (EST) "Latvia (LAT)" "Lit… "Rus… "Bel… "Arm… "Aze…  
## 10 2024 Estonia (EST) "Latvia (LAT)" "Lit… "Ind… "Ind… "Arm… "Aze…  
## [1] "Table 33"  
## # A tibble: 8 × 16  
## `` `Summer Games` `Summer Games` `Summer Games` `Summer Games`  
## <chr> <chr> <int> <int> <int>  
## 1 Team (IOC code) No. NA NA NA  
## 2 Russia (RUS) 6 147 125 150  
## 3 Russian Empire (R… 3 1 4 3  
## 4 Soviet Union (URS) 9 395 319 296  
## 5 Unified Team (EUN) 1 45 38 29  
## 6 Olympic Athletes … 0 0 0 0  
## 7 Russian Olympic C… 1 20 28 23  
## 8 Total 20 608 514 501  
## # ℹ 11 more variables: `Summer Games` <chr>, `Winter Games` <chr>,  
## # `Winter Games` <chr>, `Winter Games` <chr>, `Winter Games` <chr>,  
## # `Winter Games` <chr>, `Combined total` <chr>, `Combined total` <chr>,  
## # `Combined total` <chr>, `Combined total` <chr>, `Combined total` <chr>  
## [1] "Table 34"  
## # A tibble: 16 × 16  
## `` `Summer Games` `Summer Games` `Summer Games` `Summer Games`  
## <chr> <chr> <int> <int> <int>  
## 1 Team (IOC code) No. NA NA NA  
## 2 Estonia (EST) 14 10 9 17  
## 3 Latvia (LAT) 13 4 11 6  
## 4 Lithuania (LTU) 11 6 9 15  
## 5 Armenia (ARM) 8 2 11 9  
## 6 Belarus (BLR) 7 13 30 42  
## 7 Georgia (GEO) 8 13 15 19  
## 8 Kazakhstan (KAZ) 8 15 25 38  
## 9 Kyrgyzstan (KGZ) 8 0 5 8  
## 10 Moldova (MDA) 8 0 3 7  
## 11 Ukraine (UKR) 8 38 41 72  
## 12 Uzbekistan (UZB) 8 18 8 23  
## 13 Azerbaijan (AZE) 8 9 16 31  
## 14 Tajikistan (TJK) 8 1 1 5  
## 15 Turkmenistan (TK… 8 0 1 0  
## 16 Total 14 129 185 292  
## # ℹ 11 more variables: `Summer Games` <chr>, `Winter Games` <chr>,  
## # `Winter Games` <chr>, `Winter Games` <chr>, `Winter Games` <chr>,  
## # `Winter Games` <chr>, `Combined total` <chr>, `Combined total` <chr>,  
## # `Combined total` <chr>, `Combined total` <chr>, `Combined total` <chr>  
## [1] "Table 35"  
## # A tibble: 9 × 8  
## Date Team Team Team Team Team Team Team   
## <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr>  
## 1 1912 as part of  Austria (AUT) as part o… "" "" Serb… Serb… ""   
## 2 1920–1936 Kingdom of Yugoslavia (YUG) Kingdom o… "Kin… "Kin… King… King… "Kin…  
## 3 1948–1988 SFR Yugoslavia (YUG) SFR Yugos… "SFR… "SFR… SFR … SFR … "SFR…  
## 4 1992 W Croatia (CRO) Slovenia … "SFR… "SFR… SFR … SFR … "SFR…  
## 5 1992 S Croatia (CRO) Slovenia … "Bos… "Ind… Inde… Inde… "Ind…  
## 6 1994 Croatia (CRO) Slovenia … "Bos… "ban… ban … ban … "ban…  
## 7 1996–2006 Croatia (CRO) Slovenia … "Bos… "Nor… FR Y… FR Y… "FR …  
## 8 2008–2014 Croatia (CRO) Slovenia … "Bos… "Nor… Serb… Serb… "Mon…  
## 9 2016– Croatia (CRO) Slovenia … "Bos… "Nor… Serb… Koso… "Mon…  
## [1] "Table 36"  
## # A tibble: 12 × 16  
## `` `Summer Games` `Summer Games` `Summer Games` `Summer Games`  
## <chr> <chr> <int> <int> <int>  
## 1 Team (IOC code) No. NA NA NA  
## 2 Serbia (SRB) (19… 6 9 8 12  
## 3 Yugoslavia (YUG)… 16 26 29 28  
## 4 Independent Olym… 1 0 1 2  
## 5 Serbia and Monte… 3 2 4 3  
## 6 Croatia (CRO) (1… 9 16 15 17  
## 7 Slovenia (SLO) (… 9 10 10 11  
## 8 Bosnia and Herze… 9 0 0 0  
## 9 North Macedonia … 8 0 1 1  
## 10 Montenegro (MNE)… 5 0 1 0  
## 11 Kosovo (KOS) (20… 3 3 1 1  
## 12 Total 26 66 70 75  
## # ℹ 11 more variables: `Summer Games` <chr>, `Winter Games` <chr>,  
## # `Winter Games` <chr>, `Winter Games` <chr>, `Winter Games` <chr>,  
## # `Winter Games` <chr>, `Combined total` <chr>, `Combined total` <chr>,  
## # `Combined total` <chr>, `Combined total` <chr>, `Combined total` <chr>

# Extract the Winter Olympics table  
winter\_olympics\_table <- all\_tables[[12]]  
  
# Rename columns for the Winter Olympics table  
colnames(winter\_olympics\_table) <- c("Rank", "NOC", "Gold", "Silver", "Bronze", "Total")  
  
# Save the cleaned table to a CSV  
write\_csv(winter\_olympics\_table, "winter\_olympics\_medal\_table.csv")  
  
# Display the table  
print(winter\_olympics\_table)

## # A tibble: 48 × 6  
## Rank NOC Gold Silver Bronze Total  
## <chr> <chr> <chr> <chr> <chr> <chr>  
## 1 1 Norway 148 134 123 405   
## 2 2 United States 114 121 95 330   
## 3 3 Germany 104 98 65 267   
## 4 4 Soviet Union\* 78 57 59 194   
## 5 5 Canada 77 72 76 225   
## 6 6 Austria 71 88 91 250   
## 7 7 Sweden 65 51 60 176   
## 8 8 Switzerland 63 47 58 168   
## 9 9 Netherlands 53 49 45 147   
## 10 10 Russia 47 39 35 121   
## # ℹ 38 more rows

# Extract the summer Olympics table   
summer\_olympics\_table <- all\_tables[[11]]  
  
# Rename columns for the summer Olympics table  
colnames(summer\_olympics\_table) <- c("Rank", "NOC", "Gold", "Silver", "Bronze", "Total")  
  
# Save the cleaned table to a CSV  
write\_csv(summer\_olympics\_table, "summer\_olympics\_medal\_table.csv")  
  
# Display the table  
print(summer\_olympics\_table)

## # A tibble: 161 × 6  
## Rank NOC Gold Silver Bronze Total  
## <chr> <chr> <chr> <chr> <chr> <chr>  
## 1 1 United States 1,105 879 781 2,765  
## 2 2 Soviet Union\* 395 319 296 1,010  
## 3 3 China 303 226 198 727   
## 4 4 Great Britain 298 339 343 980   
## 5 5 France 239 278 299 816   
## 6 6 Italy 229 201 228 658   
## 7 7 Germany 218 220 255 693   
## 8 8 Japan 189 162 191 542   
## 9 9 Hungary 187 161 182 530   
## 10 10 Australia 182 192 226 600   
## # ℹ 151 more rows

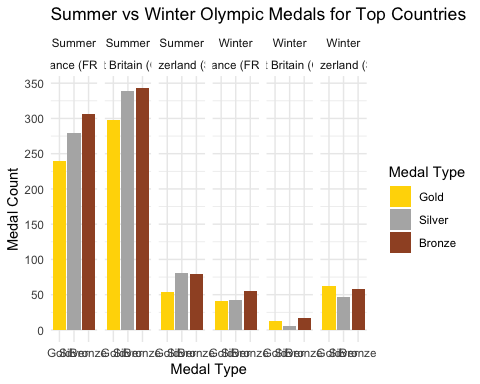
## Data Analysis

**Question 1: Which country has participated in the most Olympics (combined total), and how do its Summer and Winter Olympic medal counts (gold, silver, bronze) compare?**

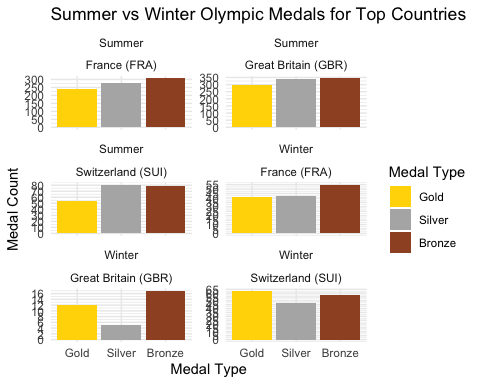
# Find all countries with the most Olympic participations, excluding the totals  
top\_countries <- first\_table\_cleaned %>%  
 mutate(Combined\_No = as.numeric(Combined\_No)) %>%  
 filter(!is.na(Team)) %>%   
 filter(Team != "Totals") %>%  
 filter(Combined\_No == max(Combined\_No, na.rm = TRUE))  
  
# Display the countries  
print(top\_countries)

## # A tibble: 3 × 16  
## Team Summer\_No Summer\_Gold Summer\_Silver Summer\_Bronze Summer\_Total Winter\_No  
## <chr> <chr> <chr> <chr> <chr> <chr> <chr>   
## 1 Fran… 30 239 279 306 821 24   
## 2 Grea… 30 298 339 343 980 24   
## 3 Swit… 30 54 81 79 214 24   
## # ℹ 9 more variables: Winter\_Gold <chr>, Winter\_Silver <chr>,  
## # Winter\_Bronze <chr>, Winter\_Total <chr>, Combined\_No <dbl>,  
## # Combined\_Gold <chr>, Combined\_Silver <chr>, Combined\_Bronze <chr>,  
## # Combined\_Total <chr>

# Reshape the data for visualization  
top\_countries\_long <- top\_countries %>%  
 select(Team, Summer\_Gold, Summer\_Silver, Summer\_Bronze, Winter\_Gold, Winter\_Silver, Winter\_Bronze) %>%  
 pivot\_longer(  
 cols = -Team,  
 names\_to = c("Season", "Medal\_Type"),  
 names\_sep = "\_",   
 values\_to = "Count"  
 ) %>%  
 mutate(  
 Count = as.numeric(Count),   
 Medal\_Type = factor(Medal\_Type, levels = c("Gold", "Silver", "Bronze"))   
 )   
  
# Bar Plot for Summer vs Winter medals  
ggplot(top\_countries\_long, aes(x = Medal\_Type, y = Count, fill = Medal\_Type)) +  
 geom\_bar(stat = "identity", position = "dodge") +  
 facet\_grid(~ Season + Team) +   
 scale\_y\_continuous(breaks = scales::pretty\_breaks(n = 10)) +   
 scale\_fill\_manual(values = c("Gold" = "gold", "Silver" = "gray70", "Bronze" = "sienna")) +   
 labs(  
 title = "Summer vs Winter Olympic Medals for Top Countries",  
 x = "Medal Type",  
 y = "Medal Count",  
 fill = "Medal Type"  
 ) +  
 theme\_minimal()



# Bar Plot for Summer vs Winter medals Version 2  
ggplot(top\_countries\_long, aes(x = Medal\_Type, y = Count, fill = Medal\_Type)) +  
 geom\_bar(stat = "identity", position = "dodge") +  
 facet\_wrap(Season ~ Team, ncol = 2, scales = "free\_y") +   
 scale\_y\_continuous(breaks = scales::pretty\_breaks(n = 10)) +   
 scale\_fill\_manual(values = c("Gold" = "gold", "Silver" = "gray70", "Bronze" = "sienna")) + # Medal colors  
 labs(  
 title = "Summer vs Winter Olympic Medals for Top Countries",  
 x = "Medal Type",  
 y = "Medal Count",  
 fill = "Medal Type"  
 ) +  
 theme\_minimal()



**Interpretation**

I created two graphs to explore different visualizations and to achieve learning objective 15. The intepretation is based on both graphs. France, Great Britain, and Switzerland share the distinction of having the most Olympic participations, each competing in 54 Olympic Games (combined Summer and Winter). Despite their similar participation records, their performances in terms of medals differ significantly. Great Britain stands out as the strongest performer, particularly in the Summer Olympics, where it has amassed 298 gold, 339 silver, and 343 bronze medals.

France follows closely behind Great Britain in terms of medal counts. In the Summer Olympics, France has earned 239 gold, 278 silver, and 299 bronze medals. While it lags behind Great Britain in overall medal tallies, France still demonstrates a strong performance and remains a major contender in the Summer Games. Switzerland, on the other hand, is the weakest performer among the three. Although Switzerland’s performance is relatively balanced between Summer and Winter Olympics, its total medal count is significantly lower than that of both France and Great Britain.

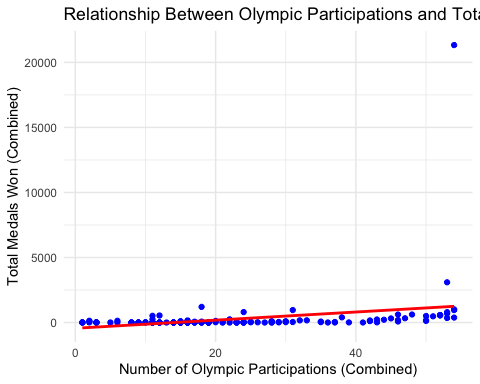
All three countries perform notably worse in the Winter Olympics compared to the Summer Olympics. This disparity can be attributed to several factors. First, there have been only 24 Winter Olympics compared to 30 Summer Olympics, which limits the opportunities to earn medals in winter events. Additionally, the Winter Olympics feature fewer sports and events than the Summer Games, resulting in fewer overall medal opportunities. Moreover, nations like France and Great Britain have historically specialized in Summer Olympic sports, such as athletics, swimming, and cycling, which are not represented in the Winter Olympics.

Geographic and climatic factors also play a role in the medal disparities. While Switzerland, with its mountainous terrain, performs relatively better in Winter Olympic sports, France and Great Britain do not have the same geographical advantages, which may limit their success in winter disciplines. Lastly, the Winter Olympics tend to attract less global investment and athlete participation compared to the Summer Olympics. This lack of focus on winter sports likely contributes to the weaker overall performance in the Winter Olympics for all three nations.

**Question 2: Is there a relationship between the number of participations and the total medals won?**

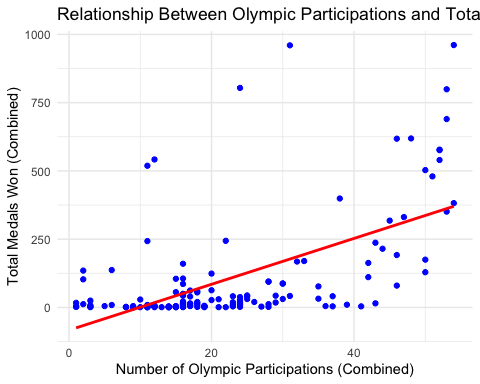
# Clean and convert Combined\_No and Combined\_Total to numeric  
first\_table\_cleaned <- first\_table\_cleaned %>%  
 mutate(  
 Combined\_No = as.numeric(gsub(",", "", Combined\_No)) \* 1,   
 Combined\_Total = as.numeric(gsub(",", "", Combined\_Total)) \* 1   
 )  
  
# Scatter plot with trend line  
ggplot(first\_table\_cleaned, aes(x = Combined\_No, y = Combined\_Total)) +  
 geom\_point(color = "blue") +   
 geom\_smooth(method = "lm", color = "red", se = FALSE) +   
 labs(  
 title = "Relationship Between Olympic Participations and Total Medals",  
 x = "Number of Olympic Participations (Combined)",  
 y = "Total Medals Won (Combined)"  
 ) +  
 theme\_minimal()

## `geom\_smooth()` using formula = 'y ~ x'



# Second graph: Filter out the outlier based on Combined\_Total  
filtered\_table <- first\_table\_cleaned %>%  
 filter((Combined\_Total \* 1) < 1000)   
  
# Scatter plot with trend line (without outlier)  
ggplot(filtered\_table, aes(x = Combined\_No, y = Combined\_Total)) +  
 geom\_point(color = "blue") +   
 geom\_smooth(method = "lm", color = "red", se = FALSE) +   
 labs(  
 title = "Relationship Between Olympic Participations and Total Medals (Without Outliers)",  
 x = "Number of Olympic Participations (Combined)",  
 y = "Total Medals Won (Combined)"  
 ) +  
 theme\_minimal()

## `geom\_smooth()` using formula = 'y ~ x'



**Interpretation**

The scatter plot demonstrates the relationship between the number of Olympic participations and the total number of medals won by countries across both Summer and Winter Olympics. There is a clear positive correlation between these two variables, as countries with more participations tend to win more medals. This trend is illustrated by the upward slope of the red regression line, which indicates that increased participation is generally associated with a higher medal count.

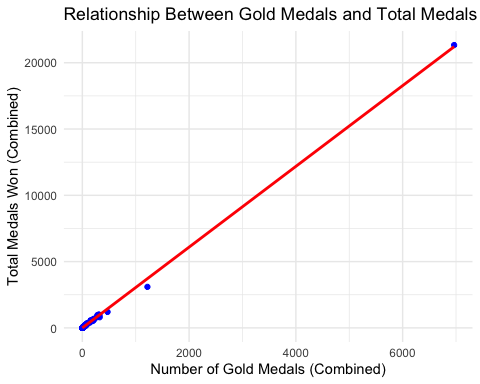
However, there is significant variability among countries. Some countries, such as the United States and the Soviet Union, are clear outliers, achieving exceptional medal counts compared to others with similar or fewer participations. These outliers highlight the impact of other factors, such as resource allocation, sports infrastructure, and historical dominance in specific sports. Conversely, several countries participate frequently but win relatively few medals, suggesting limited competitive success or a lack of specialization in medal-rich sports.

Overall, while the number of participations is an important predictor of medal counts, it is not the sole determinant of success. Historical dominance, economic investment in sports, and athlete development programs play critical roles in shaping a country’s performance. This analysis underscores the complexity of factors influencing Olympic success, beyond mere participation frequency.

**Question 3: Is there a relationship between the number of gold medals won and the total number of medals won?**

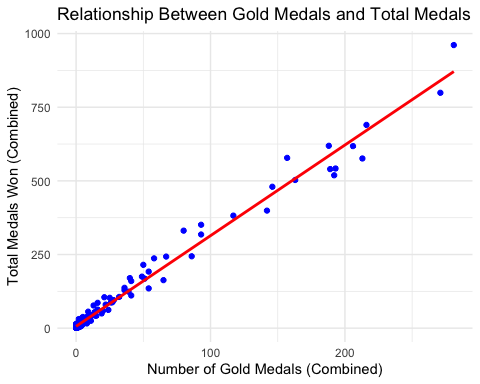
# Clean and convert Combined\_Gold and Combined\_Total to numeric  
first\_table\_cleaned <- first\_table\_cleaned %>%  
 mutate(  
 Combined\_Gold = as.numeric(gsub(",", "", Combined\_Gold)),   
 Combined\_Total = as.numeric(gsub(",", "", Combined\_Total))  
 )  
  
# Scatter plot with trend line  
ggplot(first\_table\_cleaned, aes(x = Combined\_Gold, y = Combined\_Total)) +  
 geom\_point(color = "blue") +   
 geom\_smooth(method = "lm", color = "red", se = FALSE) +   
 labs(  
 title = "Relationship Between Gold Medals and Total Medals",  
 x = "Number of Gold Medals (Combined)",  
 y = "Total Medals Won (Combined)"  
 ) +  
 theme\_minimal()

## `geom\_smooth()` using formula = 'y ~ x'



# Second graph: Filter out the outlier based on Combined\_Gold  
filtered\_table <- first\_table\_cleaned %>%  
 filter(Combined\_Gold < 300)   
  
# Scatter plot with trend line (without outlier)  
ggplot(filtered\_table, aes(x = Combined\_Gold, y = Combined\_Total)) +  
 geom\_point(color = "blue") +   
 geom\_smooth(method = "lm", color = "red", se = FALSE) +   
 labs(  
 title = "Relationship Between Gold Medals and Total Medals",  
 x = "Number of Gold Medals (Combined)",  
 y = "Total Medals Won (Combined)"  
 ) +  
 theme\_minimal()

## `geom\_smooth()` using formula = 'y ~ x'



**Interpretation**

The scatter plot explores the relationship between the number of gold medals won and the total medal count achieved by countries across the Summer and Winter Olympics. The analysis reveals a strong positive correlation, as demonstrated by the linear regression line. Countries with a higher number of gold medals typically exhibit higher total medal counts, indicating that gold-medal performance often coincides with overall success in the Olympics.

Despite this general trend, there are notable variations among countries. Some nations, such as the United States, stand out as exceptional performers, securing a disproportionate number of total medals relative to their gold medal count. This observation highlights their ability to achieve a balanced performance across all medal types (gold, silver, and bronze), suggesting a diverse and competitive athletic portfolio. On the other hand, countries with fewer total medals may have achieved fewer golds but remain competitive in silver and bronze categories, demonstrating depth in participation but not necessarily dominance.

This analysis emphasizes that while gold medal performance is a strong indicator of total Olympic success, it is not entirely deterministic. Factors such as participation in diverse events, athlete preparation, and strategic emphasis on specific sports can influence a country’s ability to consistently perform at the highest level.