"Nobel Insights: R Programming Exploration of Laureates Across Disciplines"

(2023)

Aman
Panjab University
75amansharma@gmail.com

SUMMARY OF ALL THE RELEVANT PARAMETERS

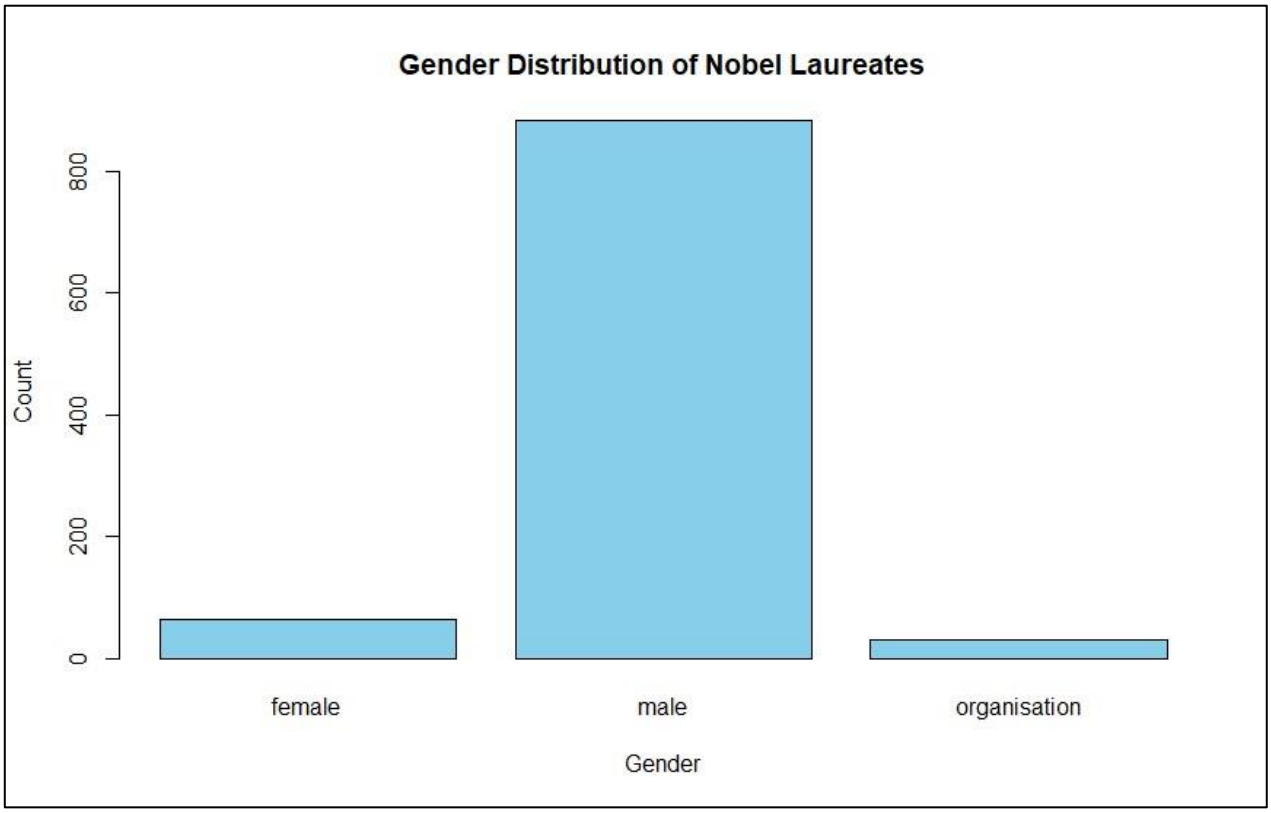
summary(nobelwinners)

```
      year        category          prizeshare      fullName            gender              born          
 Min.   :1901   Length:977         Min.   :1.000   Length:977         Length:977         Length:977        
 1st Qu.:1949   Class :character   1st Qu.:1.000   Class :character   Class :character   Class :character  
 Median :1980   Mode  :character   Median :2.000   Mode  :character   Mode  :character   Mode  :character  
 Mean   :1974                      Mean   :2.034                                                           
 3rd Qu.:2003                      3rd Qu.:3.000                                                           
 Max.   :2023                      Max.   :4.000                                                           
 bornCountry          bornCity         organizationName   organizationCountry organizationCity
 Length:977         Length:977         Length:977         Length:977          Length:977      
 Class :character   Class :character   Class :character   Class :character    Class :character
 Mode  :character   Mode  :character   Mode  :character   Mode  :character    Mode  :character
```

Question 1: What is the distribution of Nobel laureates by gender in the dataset?

```
gender_distribution <- table(nobelwinners$gender)
> gender_distribution_plot <- barplot(gender_distribution, main="
Gender Distribution of Nobel Laureates", xlab="Gender", ylab="C
ount", col="skyblue")
> gender_distribution_names <- names(gender_distribution)
```

Values	
gender_distribution	'table' int [1:3(1d)] 64 883 30
gender_distribution_names	chr [1:3] "female" "male" "organisation"



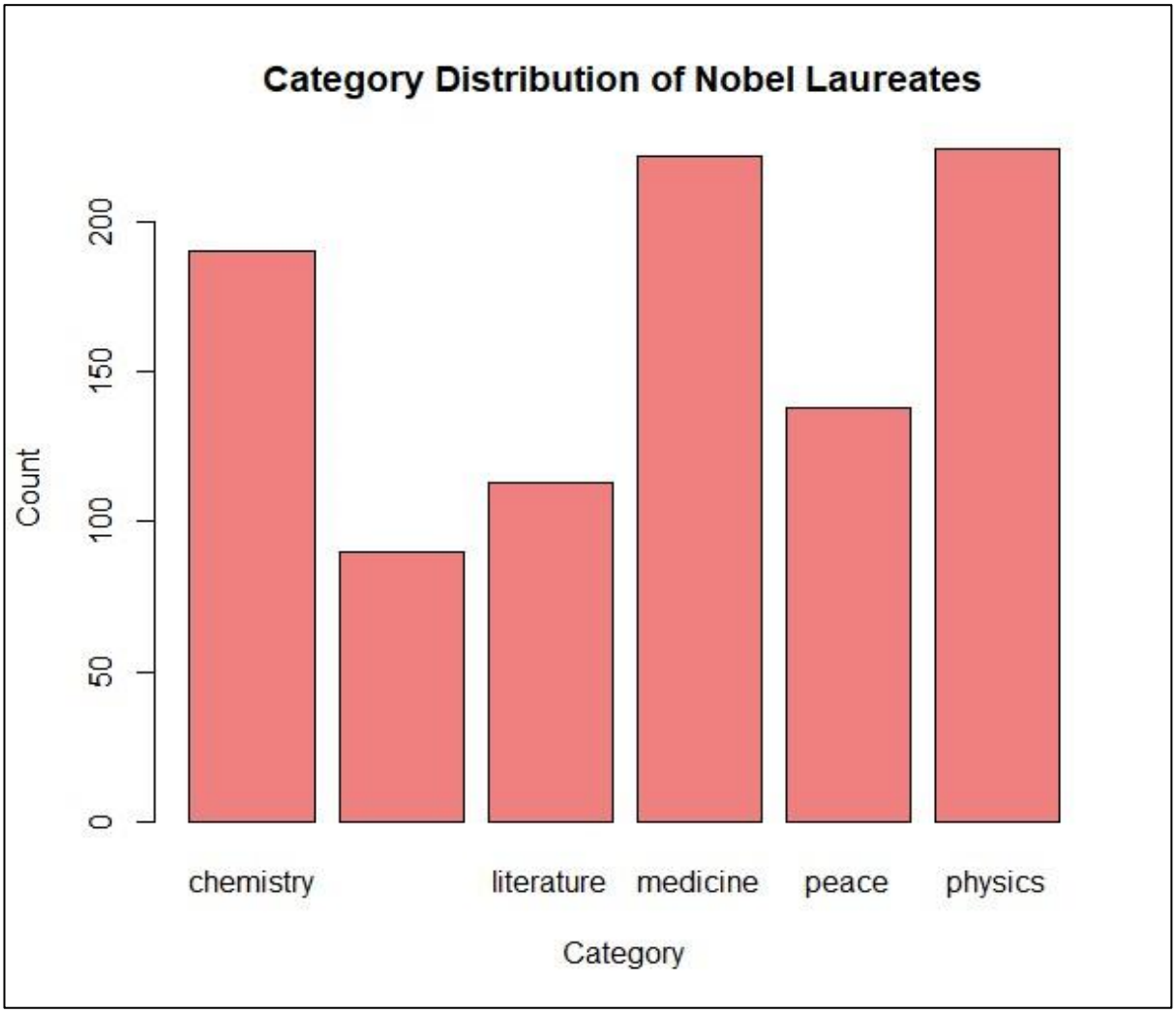
Conclusion: We can see that number of male winners are way higher than female winners, standing at 883 and 64 respectively.

Question 2: What is the distribution of Nobel laureates by category of their win in the dataset?

```
> category_distribution <- table(nobelwinners$category)
> category_distribution_plot <- barplot(category_distribution, mai
n="Category Distribution of Nobel Laureates", xlab="Category", yl
ab="Count", col="lightcoral")
> category_distribution_names <- names(category_distribution

category_distribution
```

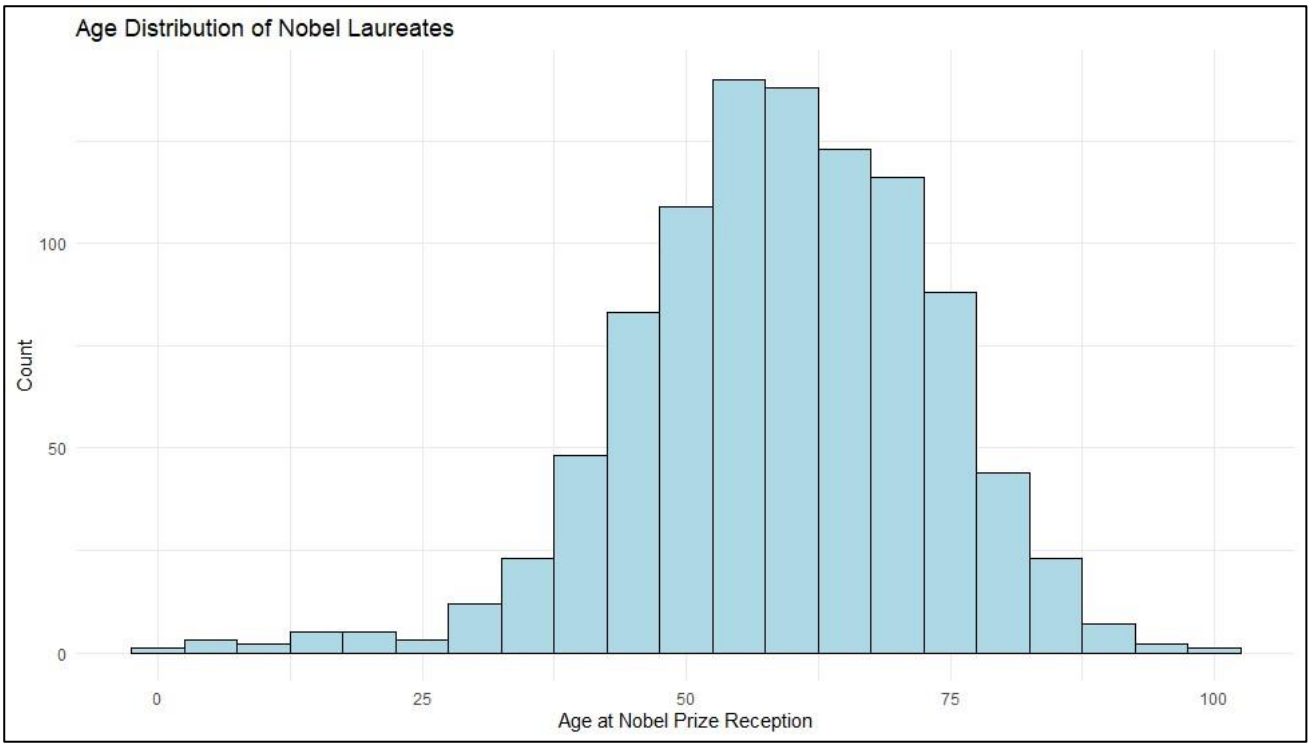
chemistry	economics	literature	medicine	peace	physics
190	90	113	222	138	224



Conclusion: We can see that majority of winners belong from Medicine (222) or Physics (224), followed by Chemistry (190). The lowest number of winners among the 6 categories is Economics (90).

Question 3: What is the distribution of Nobel laureates' ages at the time they received the Nobel Prize?

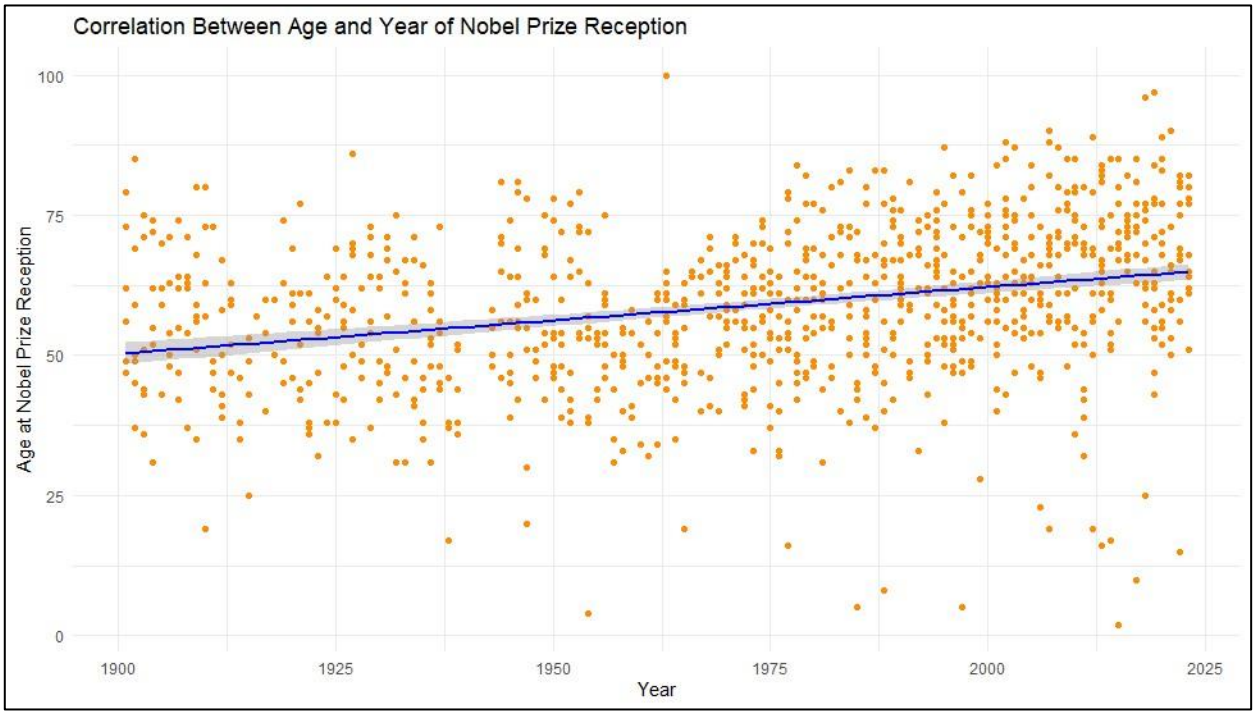
```
library(ggplot2)
> age_distribution_plot <- ggplot(nobelwinners, aes(x = age_recei
ved)) +
+    geom_histogram(binwidth = 5, fill = "lightblue", color = "black"
) +
+    labs(title = "Age Distribution of Nobel Laureates",
+         x = "Age at Nobel Prize Reception",
+         y = "Count") +
+    theme_minimal()
> age_distribution_plot
```



Conclusion: From the following distribution plot, we can see that majority of people, won the Nobel Prize at the age of 50 or above. Most people belong in the age gap of 50-75.

Question 4: Is there a correlation between the age of Nobel laureates when they received the prize and the year they received it?

```
age_vs_year_plot <- ggplot(nobelwinners, aes(x = year, y = age_re
ceived)) +
+    geom_point(color = "darkorange") +
+    geom_smooth(method = "lm", color = "blue") +
+    labs(title = "Correlation Between Age and Year of Nobel Prize
Reception",
+         x = "Year",
+         y = "Age at Nobel Prize Reception") +
+    theme_minimal()
> age_vs_year_plot
```




```
> correlation_result <- cor(nobelwinners$year, nobelwinners$age
_received)
```

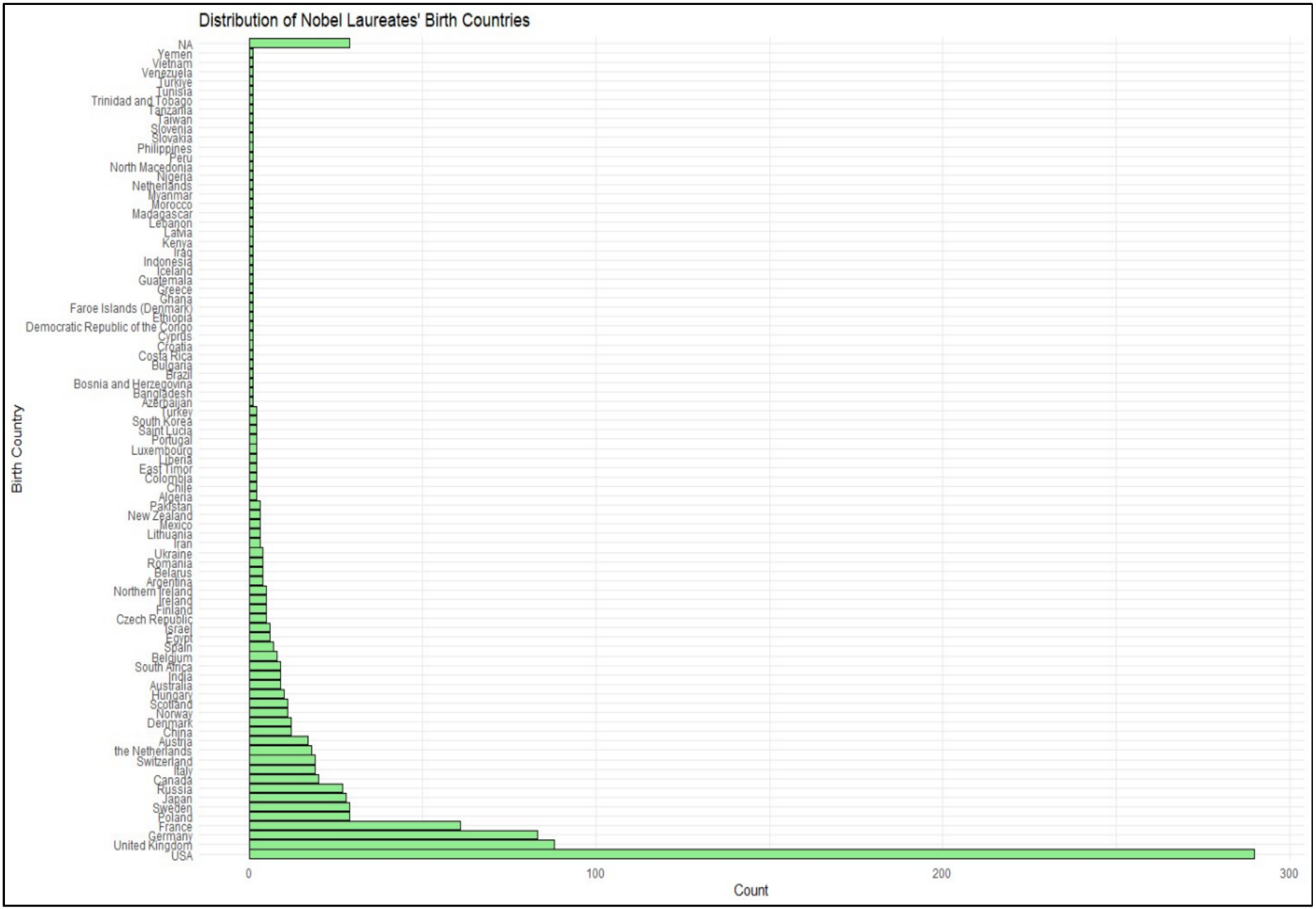
```
> print(paste("Correlation between Year and Age:", round(correlat
ion_result, 3)))
```

```
[1] "Correlation between Year and Age: 0.297"
```

Conclusion: A correlation coefficient of 0.297 is considered as low positive correlation, so we can say there exists low but a positive linear relationship between the age of Nobel laureates when they received the prize and the year they received it.

Question 5: What is the distribution of Nobel laureates' birth countries?

```
library(dplyr)
>
> birth_country_distribution <- nobelwinners %>%
+   group_by(bornCountry) %>%
+   summarize(count = n()) %>%
+   arrange(desc(count))
> birth_country_distribution_plot <- ggplot(birth_country_distribut
ion, aes(x = reorder(bornCountry, -count), y = count)) +
+   geom_bar(stat = "identity", fill = "lightgreen", color = "black") +
+   coord_flip() +
+   labs(title = "Distribution of Nobel Laureates' Birth Countries",
+        x = "Birth Country",
+        y = "Count") +
+   theme_minimal()
>
> birth_country_distribution_plot
```


Distribution of Nobel Laureates' Birth Countries
Birth Country
NA
Yemen
Vietnam
Venezuela
Turkiye
Tunisia
Trinidad and Tobago
Tanzania
Taiwan
Slovenia
Slovakia
Philippines
Peru
North Macedonia
Nigeria
Netherlands
Myanmar
Morocco
Madagascar
Lebanon
Latvia
Kenya
Iraq
Indonesia
Iceland
Guatemala
Greece
Ghana
Faroe Islands (Denmark)
Ethiopia
Democratic Republic of the Congo
Cyprus
Croatia
Costa Rica
Bulgaria
Brazil
Bosnia and Herzegovina
Bangladesh
Azerbaijan
Turkey
South Korea
Saint Lucia
Portugal
Luxembourg
Liberia
East Timor
Colombia
Chile
Algeria
Pakistan
New Zealand
Mexico
Lithuania
Iran
Ukraine
Romania
Belarus
Argentina
Northern Ireland
Ireland
Finland
Czech Republic
Israel
Egypt
Spain
Belgium
South Africa
India
Australia
Hungary
Scotland
Norway
Denmark
China
Austria
the Netherlands
Switzerland
Italy
Canada
Russia
Japan
Sweden
Poland
France
Germany
United Kingdom
USA
0
100
200
300
Count

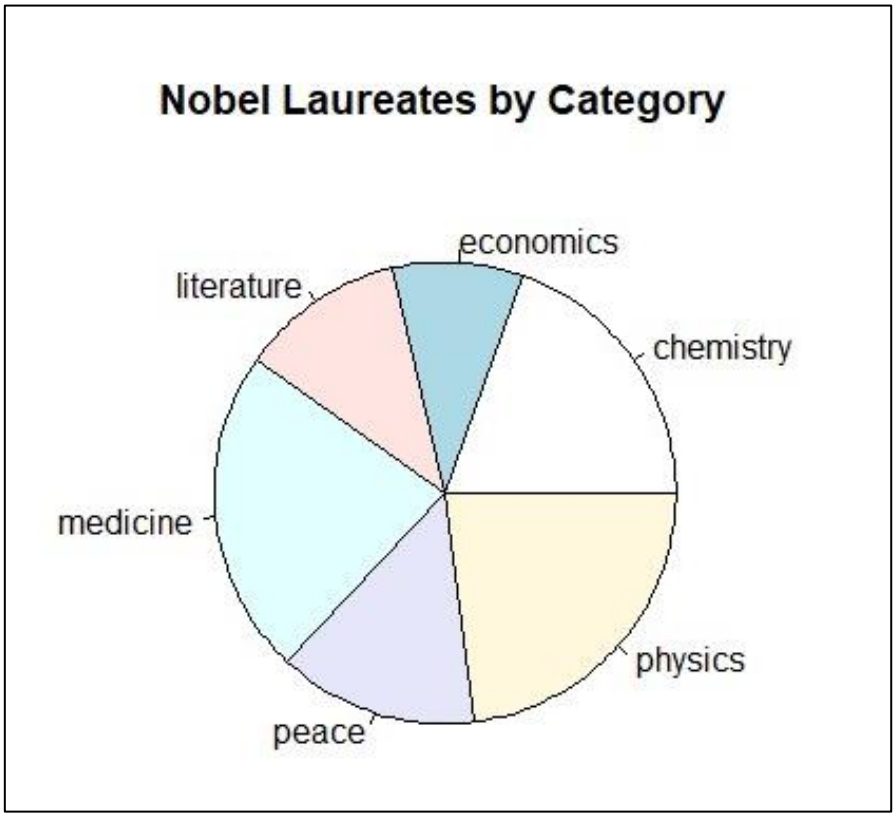
```
> country_counts <- table(nobelwinners$bornCountry)
> print(country_counts)
```

```
                Algeria                Argentina                Australia
                      2                        4                        9
                Austria                Azerbaijan                Bangladesh
                     17                        1                        1
                Belarus                  Belgium Bosnia and Herzegovina
                      4                        8                        1
                 Brazil                 Bulgaria                   Canada
                      1                        1                       20
                  Chile                    China                 Colombia
                      2                       12                        2
             Costa Rica                  Croatia                   Cyprus
                      1                        1                        1
         Czech Republic Democratic Republic of the Congo                  Denmark
                      5                        1                       12
             East Timor                    Egypt                 Ethiopia
                      2                        6                        1
Faroe Islands (Denmark)                  Finland                   France
                      1                        5                       61
                Germany                    Ghana                   Greece
                     83                        1                        1
              Guatemala                  Hungary                  Iceland
                      1                       10                        1
                  India                Indonesia                     Iran
                      9                        1                        3
                   Iraq                  Ireland                   Israel
                      1                        5                        6
                  Italy                    Japan                    Kenya
                     19                       28                        1
                 Latvia                  Lebanon                  Liberia
                      1                        1                        2
              Lithuania               Luxembourg                Madagascar
                      3                        2                        1
                 Mexico                  Morocco                  Myanmar
                      3                        1                        1
            Netherlands              New Zealand                  Nigeria
                      1                        3                        1
        North Macedonia          Northern Ireland                   Norway
                      1                        5                       11
               Pakistan                     Peru                 Philippines
                      3                        1                        1
                 Poland                 Portugal                  Romania
                     29                        2                        4
                 Russia              Saint Lucia                 Scotland
                     27                        2                       11
               Slovakia                 Slovenia             South Africa
                      1                        1                        9
            South Korea                    Spain                   Sweden
                      2                        7                       29
            Switzerland                   Taiwan                 Tanzania
                     19                        1                        1
        the Netherlands      Trinidad and Tobago                  Tunisia
                     18                        1                        1
                 Turkey                  Turkiye                  Ukraine
                      2                        1                        4
         United Kingdom                      USA                Venezuela
                     88                      290                        1
                Vietnam                    Yemen
                      1                        1
```

Conclusion: It is visible that the country with the most Nobel laureates is United States of America (290), followed by United Kingdom (88).

Question 6: Create a pie chart to visualize the distribution of Nobel laureates by category.

```
> category_distribution <- table(nobelwinners$category)
> pie(category_distribution, labels = names(category_distribution
), main = "Nobel Laureates by Category")
```



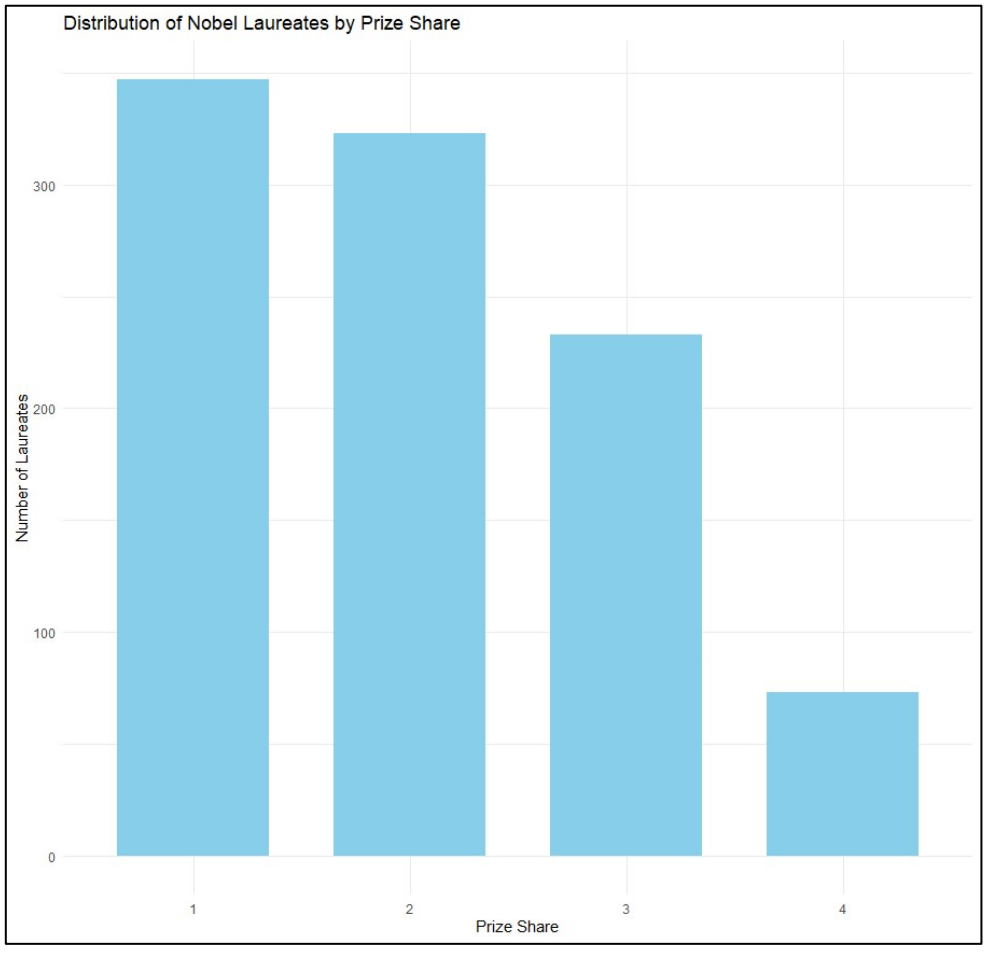
Conclusion: It can be noticed that medicine and physics contribute to the majority of the pie chart, almost 50 percent.

Question 7: How many Nobel laureates have received the prize individually or have shared the prize.

```
> prize_share_distribution <- table(nobelwinners$prizeShare)
> print(prize_share_distribution)
```

```
  1   2   3   4
347 323 233  73
```

```
> library(ggplot2)
> ggplot(nobelwinners, aes(x = factor(prizeShare))) +
+   geom_bar(stat = "count", fill = "skyblue", width = 0.7) +
+   labs(title = "Distribution of Nobel Laureates by Prize Share",
+        x = "Prize Share",
+        y = "Number of Laureates") +
+   theme_minimal()
```



Conclusion: It can be noticed that most number of the winners (347) have not shared the prize. While 629 winners have shared it with one or more than one winner.

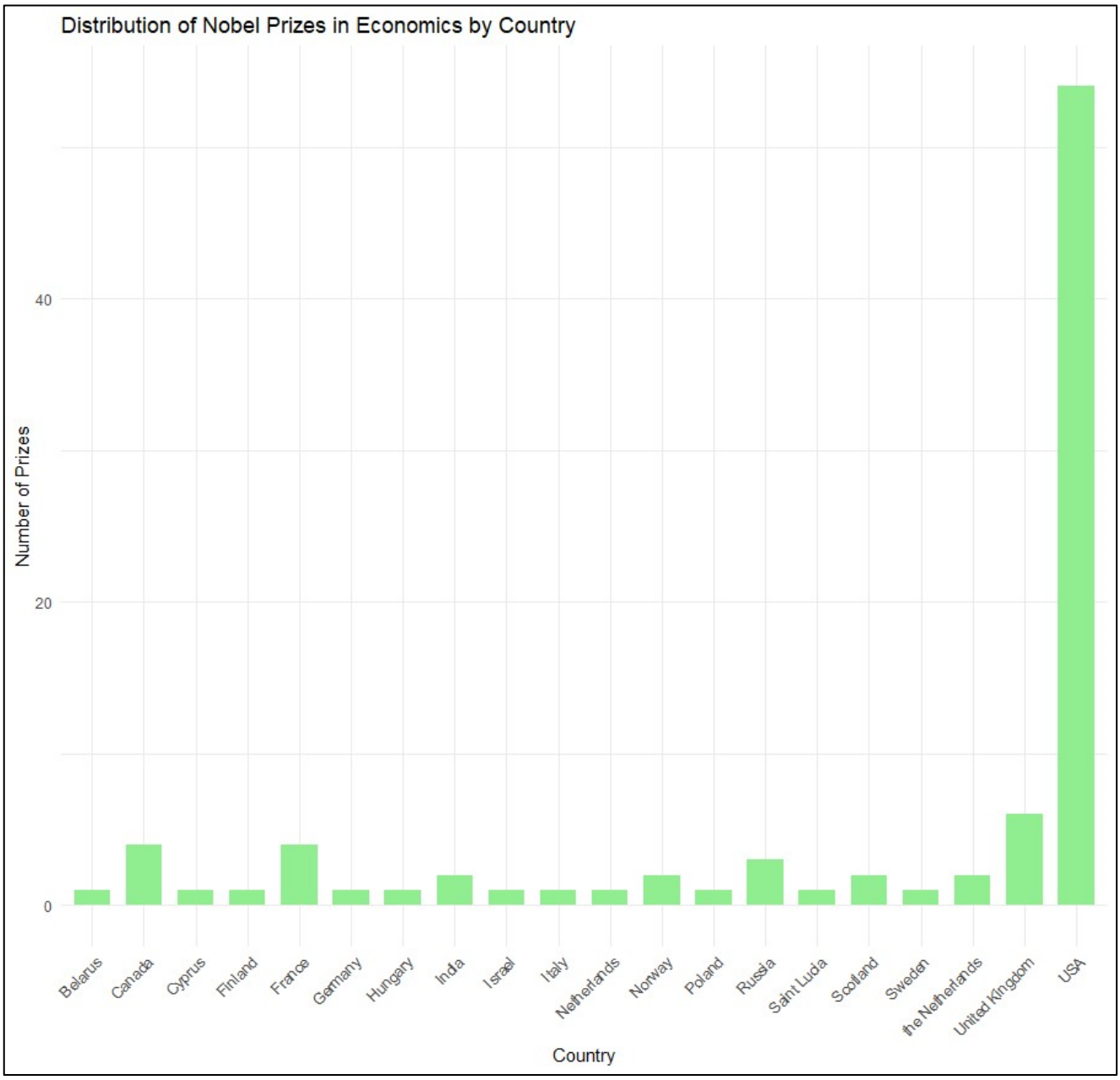
Question 8: Which country has received the most Nobel Prizes in Economics?

```
> library(dplyr)
> economics_prizes <- nobelwinners %>% filter(category == "eco
nomics")
>
> country_prizes <- economics_prizes %>% group_by(bornCountr
y) %>% summarise(total_prizes = n())
>
>
> most_prizes_country <- country_prizes[which.max(country_priz
es$total_prizes), ]
>
> print(most_prizes_country)
```

	bornCountry	total_prizes
	<chr>	<int>
1	USA	54

```
> ggplot(economics_prizes, aes(x = factor(bornCountry))) +
+    geom_bar(stat = "count", fill = "lightgreen", width = 0.7) +
+    labs(title = "Distribution of Nobel Prizes in Economics by Cou
ntry",
+         x = "Country",
+         y = "Number of Prizes") +
+    theme_minimal() +
```

```
+    theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Conclusion: It can be noticed that most number of the winners (347) have not shared the prize. While 629 winners have shared it with one or more than one winner.

Question 9: How many Nobel Prizes have been received by individuals or organizations associated with India?

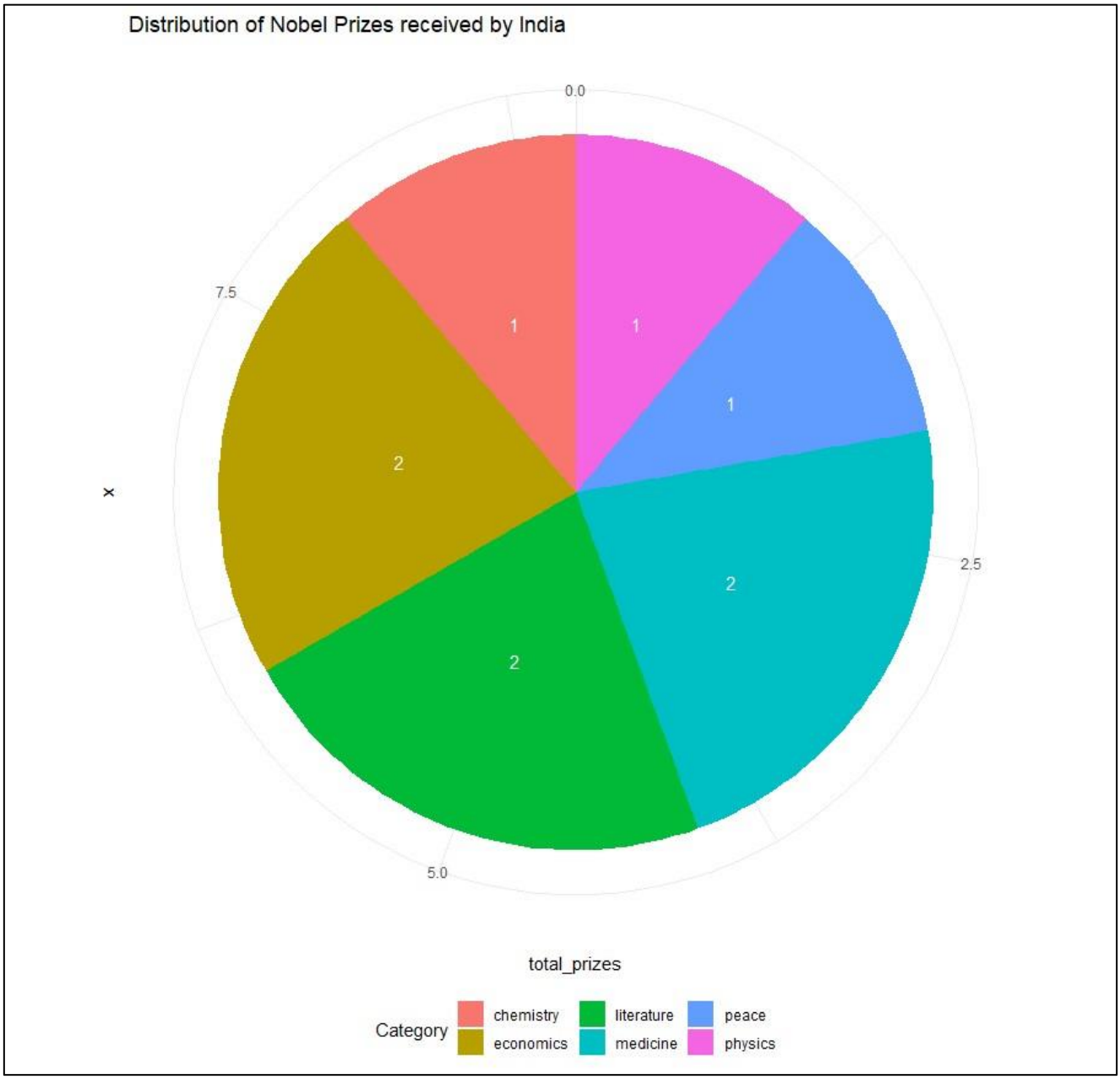
```
> india_prizes <- nobelwinners %>% filter(bornCountry == "India" | organizationCountry == "India")
>
> total_india_prizes <- nrow(india_prizes)
> print(paste("Total Nobel Prizes received by individuals or organizations associated with India:", total_india_prizes))
```

Output = "Total Nobel Prizes received by individuals or organizations associated with India: 9"

```
> category_counts <- india_prizes %>% group_by(category) %>% summarise(total_prizes = n())
> most_prizes_category <- category_counts[which.max(category_counts$total_prizes), ]
> print(paste("Category with the most Nobel Prizes received by India:", most_prizes_category$category))
```

Output = "Category with the most Nobel Prizes received by India: economics"

```
> ggplot(category_counts, aes(x = "", y = total_prizes, fill = category)) +
+    geom_bar(stat = "identity", width = 1) +
+    coord_polar("y") +
+    geom_text(aes(label = total_prizes), position = position_stack(vjust = 0.5), size = 4, color = "white") +
+    labs(title = "Distribution of Nobel Prizes received by India",
+         fill = "Category") +
+    theme_minimal() +
+    theme(legend.position = "bottom")
```



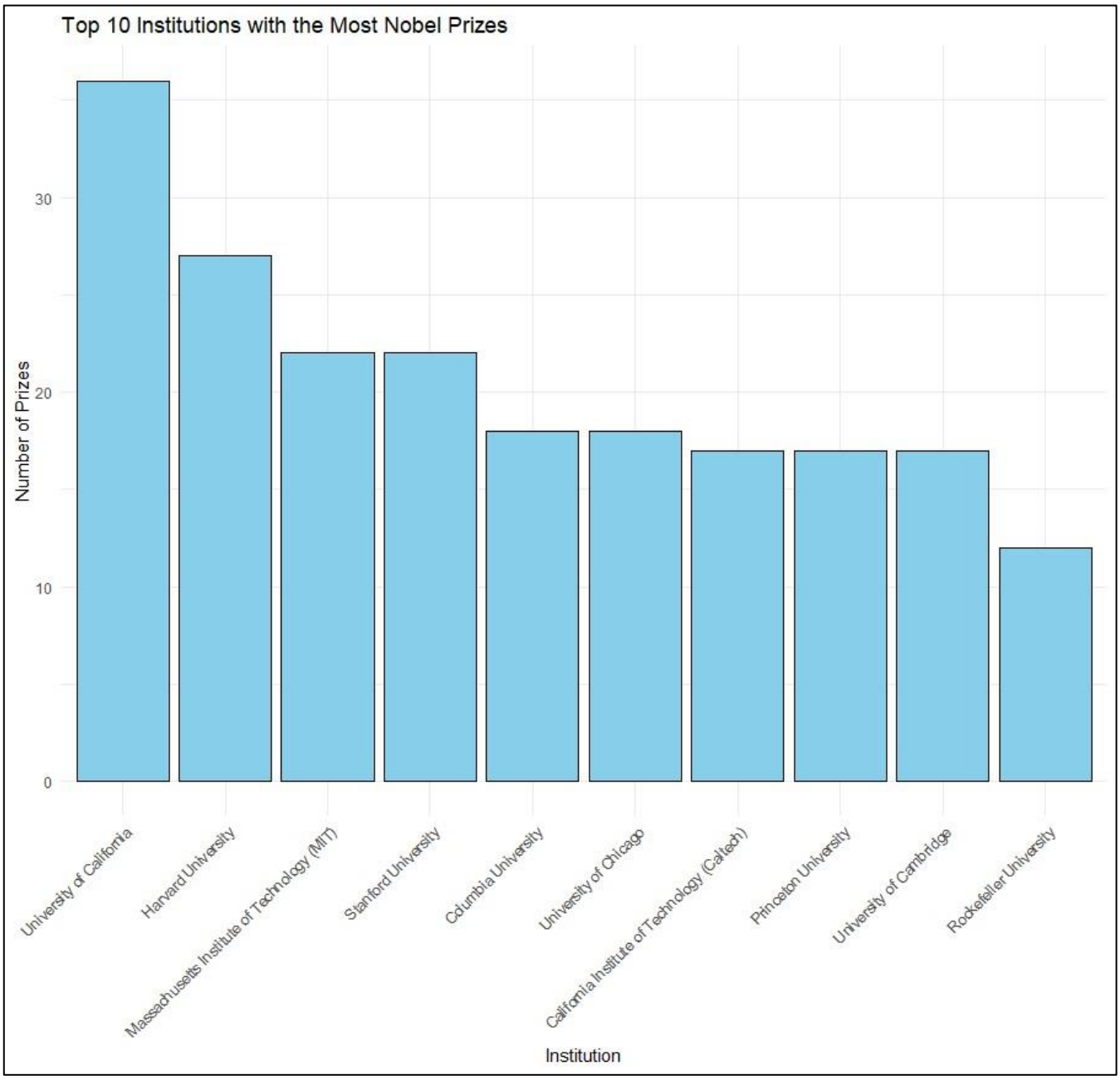
Conclusion: There is almost an even share among all the categories, with 2 prizes each from Literature, Peace, and Economics, and 1 each from Chemistry, Medicine, and Physics, making it a total of 9.

Question 10: Which institution has received the most Nobel Prizes across all categories?

```
> institution_counts <- nobelwinners %>%
+   filter(!is.na(organizationName)) %>%  # Exclude rows with mis
sing institution data
+   group_by(organizationName) %>%
+   summarise(total_prizes = n()) %>%
+   arrange(desc(total_prizes))
>
> most_prizes_institution <- institution_counts[1, ]
>
> print(paste("Institution with the most Nobel Prizes:", most_prize
s_institution$organizationName))
```

Output = "Institution with the most Nobel Prizes: University of California"

```
> top_institutions <- head(institution_counts, 10)
>
> ggplot(top_institutions, aes(x = reorder(organizationName, -tota
l_prizes), y = total_prizes)) +
+   geom_bar(stat = "identity", fill = "skyblue", color = "black") +
+   labs(title = "Top 10 Institutions with the Most Nobel Prizes",
+        x = "Institution",
+        y = "Number of Prizes") +
+   theme_minimal() +
+   theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Conclusion: The institution with the most Nobel Winners is University of California, followed by Harvard University, and MIT.

THANK YOU