Analysis of World's Top Intelligent Individuals

Hasan Sezer

2024-08-16

1. Data Preparation

```r
topIntelligent <- read.csv("top_intelligent.csv")
glimpse(topIntelligent)
```

```
## Rows: 5,000
## Columns: 11
## $ Name               <chr> "Enrico Fermi", "Max Planck", "Paul Dirac", "Erwin ~
## $ Country            <chr> "Austria", "Italy", "UK", "Italy", "UK", "USA", "Ge~
## $ Field.of.Expertise <chr> "Polymath", "Chemistry", "Physics", "Physics", "Phy~
## $ IQ                 <int> 199, 159, 177, 130, 163, 191, 187, 186, 149, 182, 1~
## $ Achievements       <chr> "Father of Computer Science", "Theory of Evolution"~
## $ Birth.Year         <int> 1968, 1986, 1927, 1921, 1964, 1990, 1966, 1908, 192~
## $ Gender             <chr> "Female", "Female", "Female", "Female", "Female", "~
## $ Notable.Works      <chr> "E=mc²", "Bohr Model", "Cosmos", "Discovery of Elec~
## $ Awards             <chr> "Numerous Posthumous", "Nobel Prize", "Nobel Prize"~
## $ Education          <chr> "Self-taught", "Ph.D. in Astronomy", "Ph.D. in Math~
## $ Influence          <chr> "Popularizing science and cosmology", "Foundational~
```

2. Data Cleaning

Checking for "N/A" values

```r
sapply(topIntelligent, function(x) sum(x == "N/A"))
```

```
##               Name            Country Field.of.Expertise                 IQ
##                  0                  0                  0                  0
##       Achievements         Birth.Year             Gender      Notable.Works
##                  0                  0                  0                  0
##             Awards          Education          Influence
##               1249                  0                  0
```

```r
table(topIntelligent$Awards)
```

```
##
##        Copley Medal                 N/A         Nobel Prize     Numerous Honors
##                 247                1249                1951                 505
## Numerous Posthumous     Pulitzer Prize         Royal Medal    Two Nobel Prizes
##                 251                 270                 274                 253
```

- Awards column has 1249 NAs in string format.
- Dropping NAs in Awards

```r
topIntelligent_cleaned <- subset(topIntelligent, Awards != "N/A")
table(topIntelligent_cleaned$Awards)
```

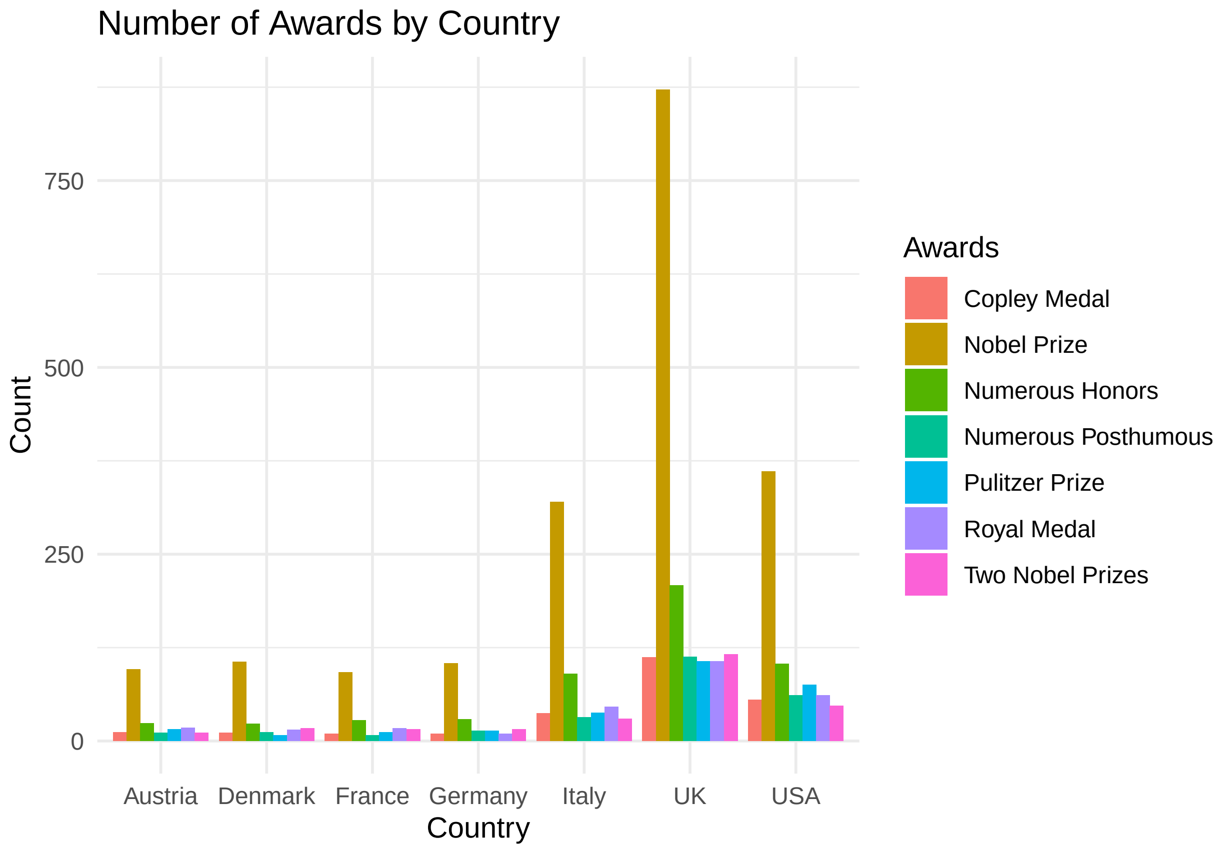
```
##
##       Copley Medal          Nobel Prize     Numerous Honors Numerous Posthumous
##                247                 1951                 505                 251
##     Pulitzer Prize          Royal Medal    Two Nobel Prizes
##                270                  274                 253
```

3. Descriptive Statistics

3.1 Distribution of Awards by Each Country

```r
Awards_by_Country <- topIntelligent_cleaned %>%
  count(Country, Awards)
```

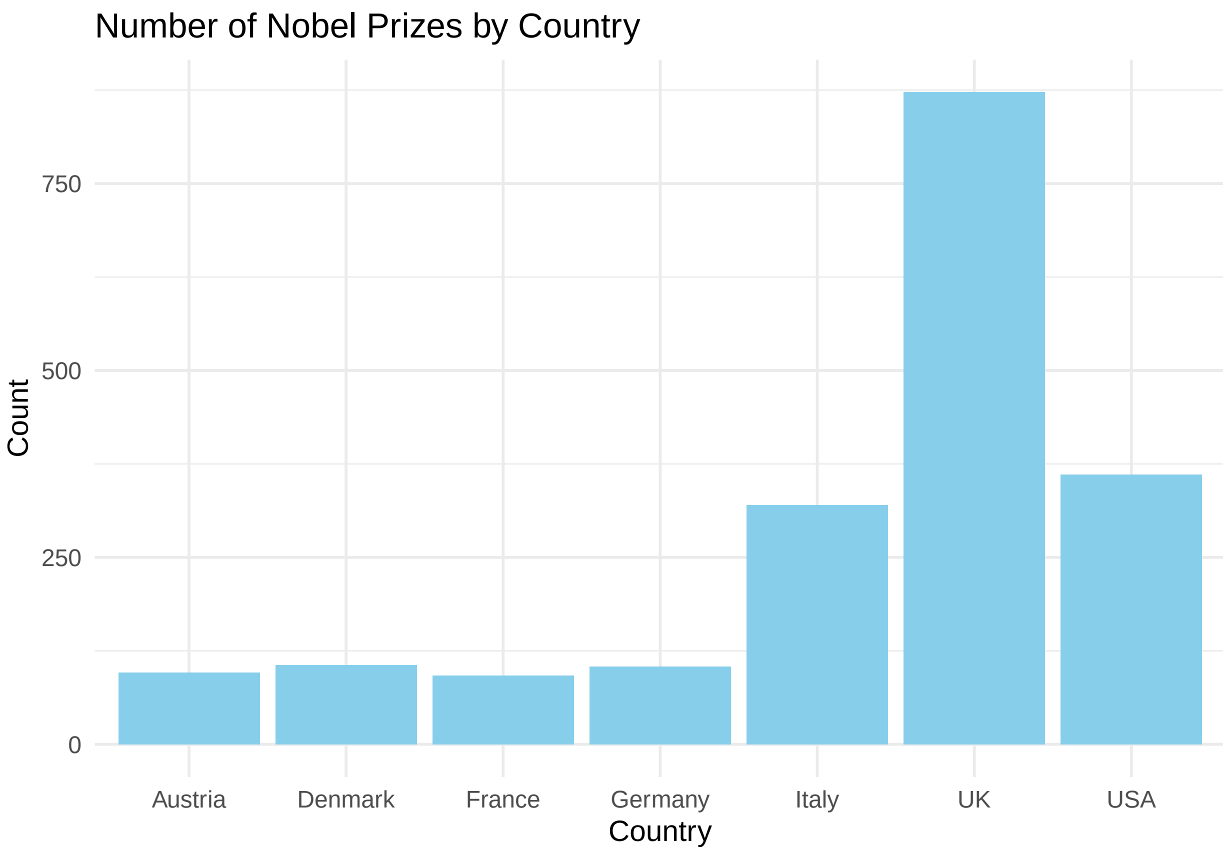
```r
ggplot(data = Awards_by_Country, aes(x = Country, y = n, fill = Awards)) +
  geom_bar(stat = "identity", position = "dodge") +
  scale_y_continuous(labels = scales::label_number()) +
  labs(title = "Number of Awards by Country",
       x = "Country",
       y = "Count") +
  theme_minimal()
```



3.2 The Country with the Most Nobel Prizes

```
NobelPrize <- Awards_by_Country %>%
  filter(Awards == "Nobel Prize")
```

```
ggplot(data = NobelPrize, aes(x = Country, y = n)) +
  geom_bar(stat = "identity", position = "dodge", fill = "Skyblue") +
  scale_y_continuous(labels = scales::label_number()) +
  labs(title = "Number of Nobel Prizes by Country",
       x = "Country",
       y = "Count") +
  theme_minimal()
```



3.3 The Country with the Most Awards

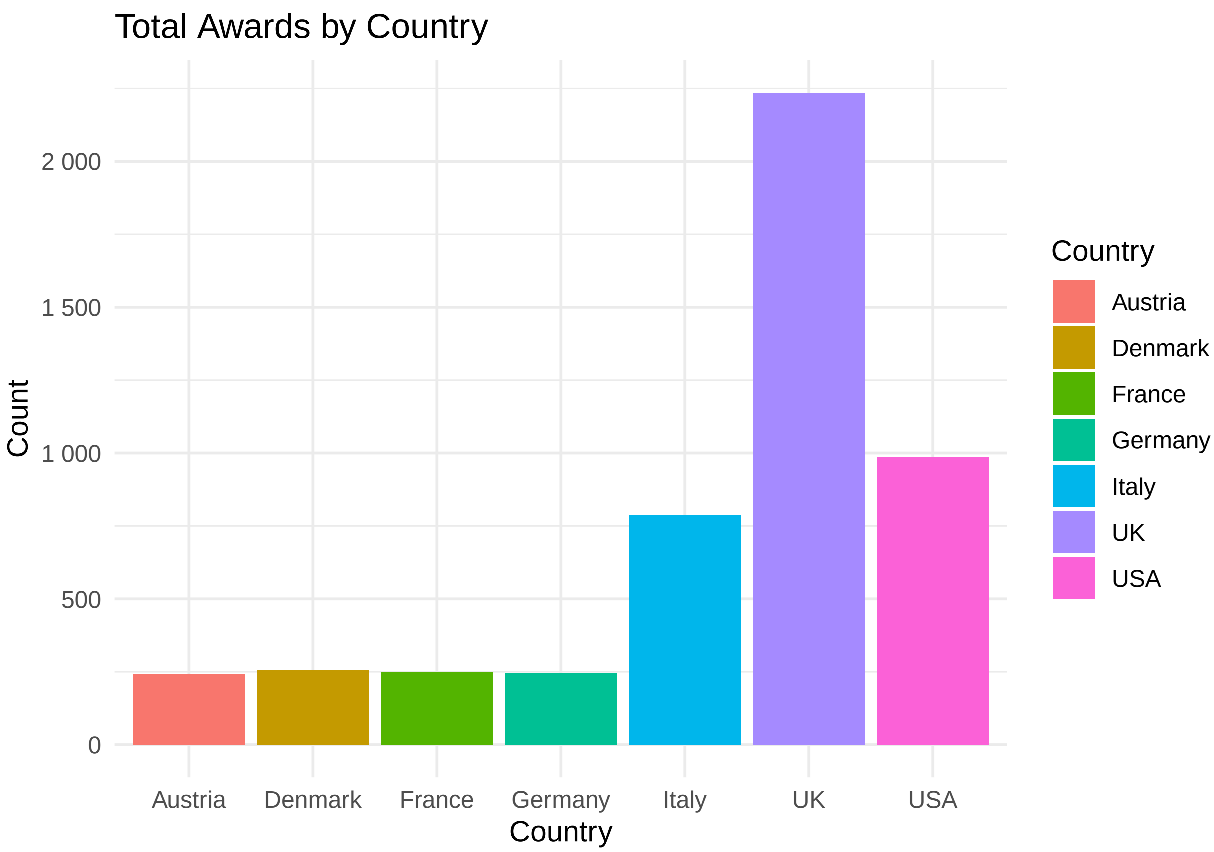
```
sum_country_awards <- topIntelligent %>%
  group_by(Country) %>%
  summarize(Total_Awards = n())

sum_country_awards %>%
  arrange(desc(Total_Awards))
```

```
## # A tibble: 7 x 2
##   Country Total_Awards
##   <chr>          <int>
## 1 UK              2235
## 2 USA              987
```

```
## 3 Italy           786
## 4 Denmark         257
## 5 France          249
## 6 Germany         245
## 7 Austria         241
```

```
ggplot(data = sum_country_awards, aes(x = Country, y = Total_Awards, fill = Country)) +
  geom_bar(stat = "identity", position = "dodge") +
  scale_y_continuous(labels = scales::label_number()) +
  labs(title = "Total Awards by Country",
       x = "Country",
       y = "Count") +
  theme_minimal()
```



3.4 IQ Analysis

```
summary(topIntelligent$IQ)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   130.0   147.0   165.0   164.8   182.0   200.0
```

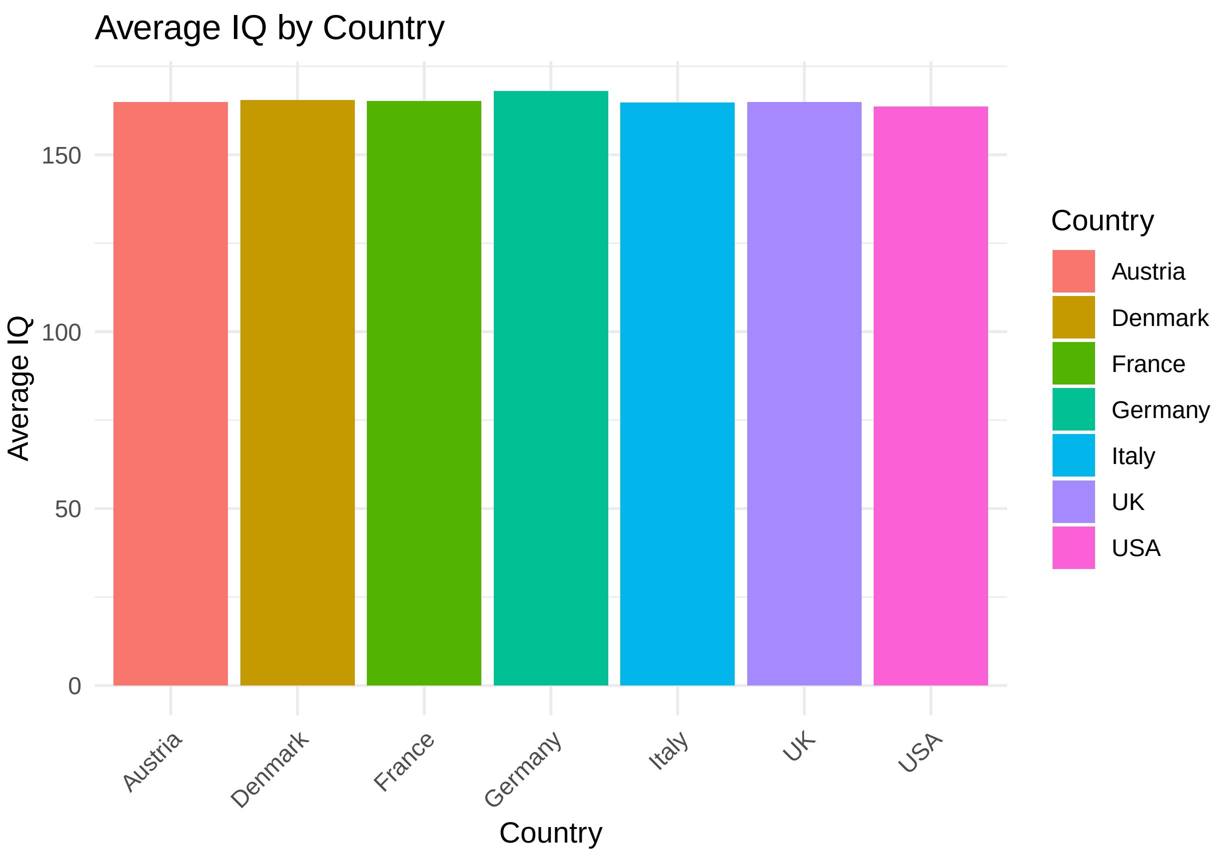
IQ and Country

```
IQ_by_Country <- topIntelligent %>%
  group_by(Country) %>%
  summarize(Average_IQ = mean(IQ, na.rm = TRUE))

IQ_by_Country
```

```
## # A tibble: 7 x 2
##   Country Average_IQ
##   <chr>        <dbl>
## 1 Austria        165.
## 2 Denmark        165.
## 3 France         165.
## 4 Germany        168.
## 5 Italy          165.
## 6 UK             165.
## 7 USA            164.
```

```
ggplot(data = IQ_by_Country, aes(x = Country, y = Average_IQ, fill = Country)) +
  geom_bar(stat = "identity") +
  scale_y_continuous(labels = scales::label_number()) +
  labs(title = "Average IQ by Country",
       x = "Country",
       y = "Average IQ") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Interpretation:

There is almost no IQ variance by country.

- IQ and Education

```
table(topIntelligent$Education)
```

```
##
##                     Ph.D.         Ph.D. (honorary)       Ph.D. in Astronomy
```

```
##                        232                         265                         266
##      Ph.D. in Chemistry    Ph.D. in Mathematics        Ph.D. in Physics
##                        531                         727                        1796
##              Self-taught University of Cambridge      University of Pisa
##                        706                         243                         234
```

```r
# IQ by Education
IQ_by_Education <- topIntelligent %>%
  group_by(Education) %>%
  summarise(Avr_IQ = mean(IQ, na.rm = TRUE))

IQ_by_Education
```

```
## # A tibble: 9 x 2
##   Education                Avr_IQ
##   <chr>                     <dbl>
## 1 Ph.D.                      166.
## 2 Ph.D. (honorary)           164.
## 3 Ph.D. in Astronomy         163.
## 4 Ph.D. in Chemistry         166.
## 5 Ph.D. in Mathematics       165.
## 6 Ph.D. in Physics           164.
## 7 Self-taught                165.
## 8 University of Cambridge    163.
## 9 University of Pisa         165.
```

- IQ and Influence

```r
table(topIntelligent$Influence)
```

```
##
## Contributions to cosmology and quantum gravity
##                                            255
##                        Developed quantum theory
##                                            251
##         Discovery of electromagnetic induction
##                                            253
##              Early pioneer of computer science
##                                            247
##                   Evolutionary biology pioneer
##                                            273
##              Foundation of classical mechanics
##                                            243
##         Foundational work in quantum mechanics
##                                            472
##         Iconic Renaissance artist and inventor
##                                            255
##     Inventor and electrical engineering pioneer
##                                            239
##       Laid the groundwork for modern computing
##                                            278
##               Major advancements in astronomy
##                                            238
##        Nuclear physics and reactor development
##                                            242
##    Penrose's work on black holes and cosmology
```

```
##                                              235
##           Pioneering research in radioactivity
##                                              247
##             Popularizing science and cosmology
##                                              253
##   Quantum electrodynamics and physics teaching
##                                              254
##                  Revolutionized modern physics
##                                              280
##             Unified theory of electromagnetism
##                                              238
##              Wave mechanics in quantum physics
##                                              247
```

```r
# IQ by Influence
IQ_by_Influence <- topIntelligent %>%
  group_by(Influence) %>%
  summarize(Average_IQ = mean(IQ, na.rm = TRUE))

IQ_by_Influence
```

```
## # A tibble: 19 x 2
##    Influence                                      Average_IQ
##    <chr>                                               <dbl>
##  1 Contributions to cosmology and quantum gravity       165.
##  2 Developed quantum theory                             165.
##  3 Discovery of electromagnetic induction               166.
##  4 Early pioneer of computer science                    163.
##  5 Evolutionary biology pioneer                         166.
##  6 Foundation of classical mechanics                    167.
##  7 Foundational work in quantum mechanics               162.
##  8 Iconic Renaissance artist and inventor               165.
##  9 Inventor and electrical engineering pioneer          165.
## 10 Laid the groundwork for modern computing             167.
## 11 Major advancements in astronomy                      164.
## 12 Nuclear physics and reactor development              165.
## 13 Penrose's work on black holes and cosmology          166.
## 14 Pioneering research in radioactivity                 165.
## 15 Popularizing science and cosmology                   164.
## 16 Quantum electrodynamics and physics teaching         166.
## 17 Revolutionized modern physics                        165.
## 18 Unified theory of electromagnetism                   163.
## 19 Wave mechanics in quantum physics                    164.
```

- IQ by Awards

```r
IQ_by_Awards <- topIntelligent_cleaned %>%
  group_by(Awards) %>%
  summarise(Avr_IQ = mean(IQ, na.rm = TRUE))

IQ_by_Awards
```

```
## # A tibble: 7 x 2
##   Awards               Avr_IQ
##   <chr>                 <dbl>
## 1 Copley Medal           165.
```

```
## 2 Nobel Prize          166.
## 3 Numerous Honors      165.
## 4 Numerous Posthumous  165.
## 5 Pulitzer Prize       162.
## 6 Royal Medal          164.
## 7 Two Nobel Prizes     165.
```

Interpretation: IQ does not show variance across the variables inspected so far.

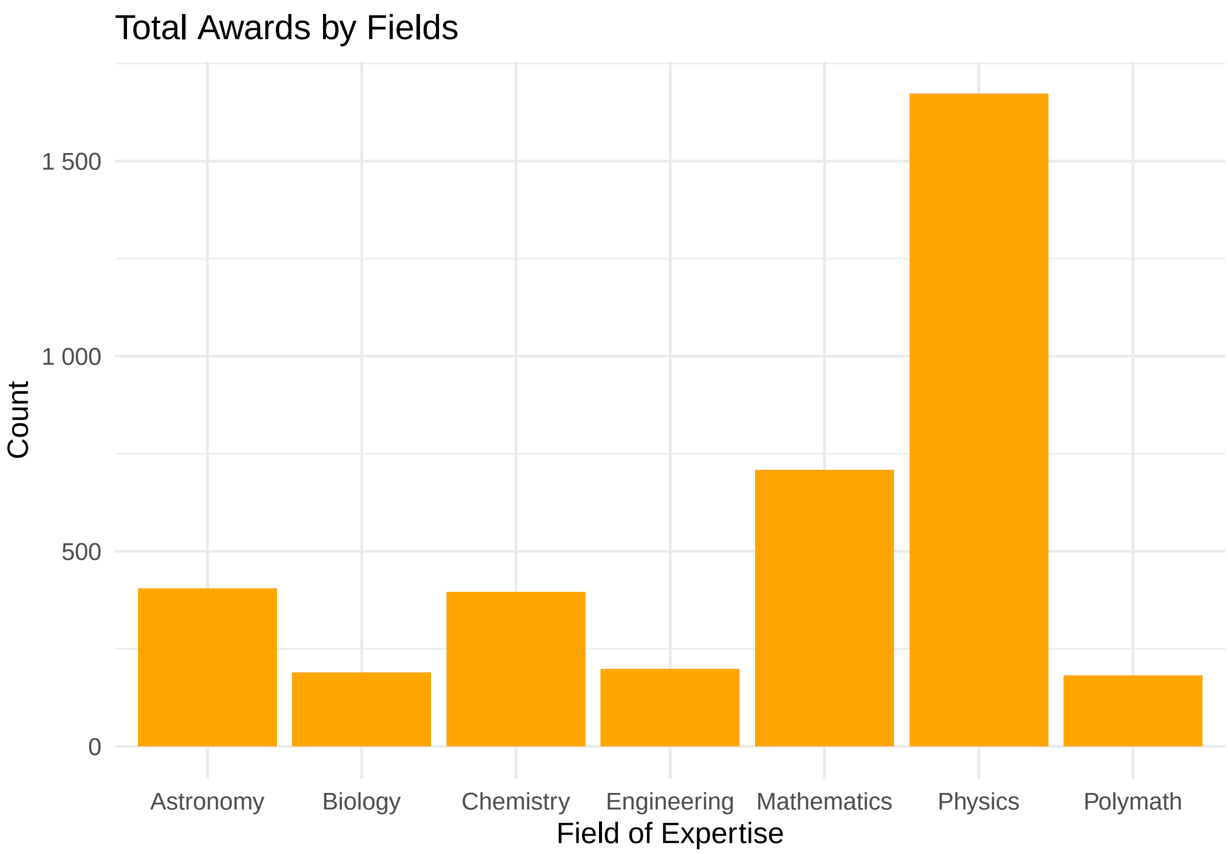
3.5 Fields Analysis

- Field with the most awards

```
award_by_field <- topIntelligent_cleaned %>%
  count(Field.of.Expertise, Awards)
```

```
field_with_most_awards <- award_by_field %>%
  group_by(Field.of.Expertise) %>%
  summarise(total_awards = sum(n))
```

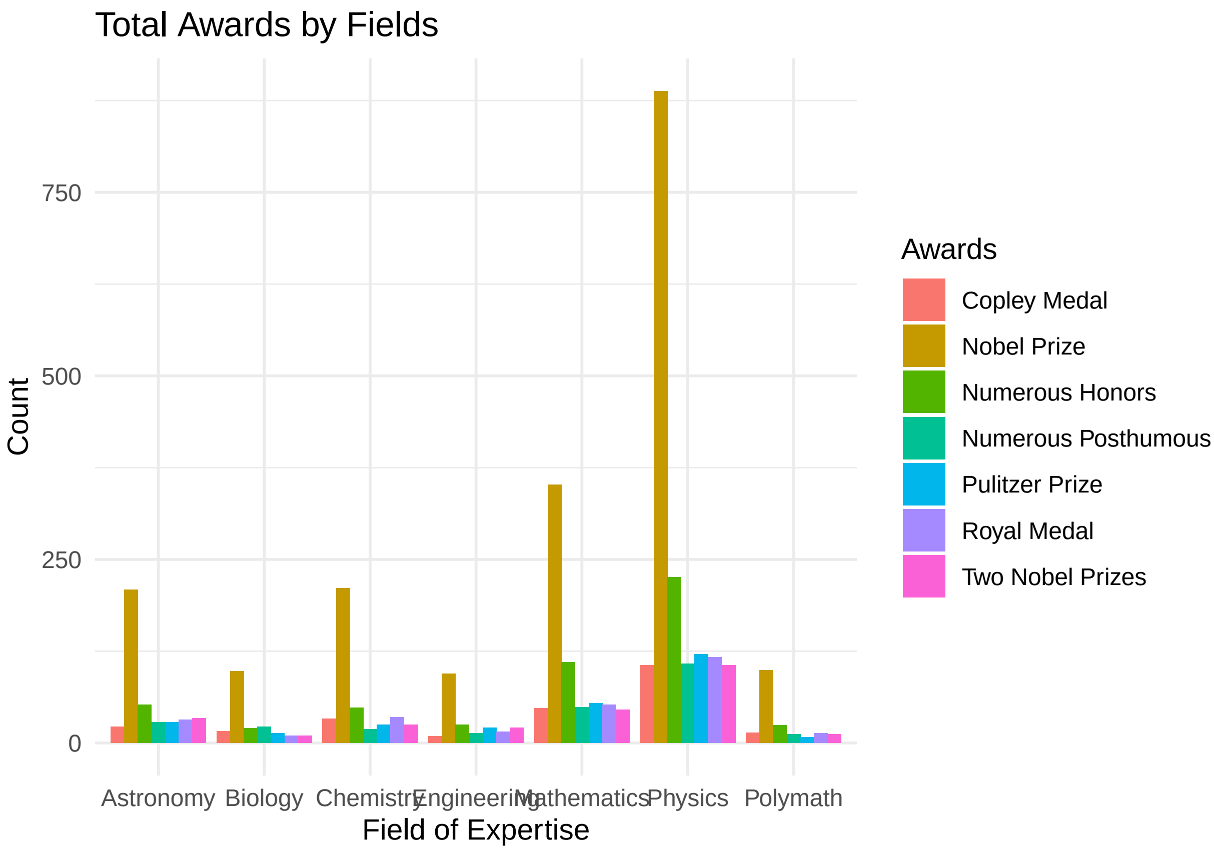
```
ggplot(data = field_with_most_awards, aes(x = Field.of.Expertise, y = total_awards)) +
  geom_bar(stat = "identity", position = "dodge", fill = "orange") +
  scale_y_continuous(labels = scales::label_number()) +
  labs(title = "Total Awards by Fields",
       x = "Field of Expertise",
       y = "Count") +
  theme_minimal()
```



- Distribution of different awards in varying fields.

```r
field_by_award <- topIntelligent_cleaned %>%
  count(Field.of.Expertise, Awards)

ggplot(data = field_by_award, aes(x = Field.of.Expertise, y = n, fill = Awards)) +
  geom_bar(stat = "identity", position = "dodge") +
  scale_y_continuous(labels = scales::label_number()) +
  labs(title = "Total Awards by Fields",
       x = "Field of Expertise",
       y = "Count") +
  theme_minimal()
```



Interpretation: Physics has the most awards. And in Physics, Nobel Prize is granted most.

4. How Can One Get a Nobel Prize: Regression Analysis

In this section, I will try to predict what factors contribute to receiving a Nobel Prize.

4.1 Logistic Regression Analysis

Creating a New Column for Nobel Prize

I will create a new column, `NobelPrize`, with dummy variables: - `0` will be the reference category for not getting a Nobel Prize. - `1` will indicate receiving a Nobel Prize.

```r
# Creating the Nobel Prize dummy variable
topIntelligent_cleaned$NobelPrize <- ifelse(topIntelligent_cleaned$Awards == "Nobel Prize", 1, 0)
```

```r
str(topIntelligent_cleaned)
```

```
## 'data.frame':    3751 obs. of  12 variables:
##  $ Name              : chr  "Enrico Fermi" "Max Planck" "Paul Dirac" "Erwin Schrödinger" ...
##  $ Country           : chr  "Austria" "Italy" "UK" "Italy" ...
##  $ Field.of.Expertise: chr  "Polymath" "Chemistry" "Physics" "Physics" ...
##  $ IQ                : int  199 159 177 130 163 191 186 160 141 185 ...
##  $ Achievements      : chr  "Father of Computer Science" "Theory of Evolution" "Quantum Mechanics" "I
##  $ Birth.Year        : int  1968 1986 1927 1921 1964 1990 1908 1926 1938 1986 ...
##  $ Gender            : chr  "Female" "Female" "Female" "Female" ...
##  $ Notable.Works     : chr  "E=mc²" "Bohr Model" "Cosmos" "Discovery of Electromagnetic Induction" .
##  $ Awards            : chr  "Numerous Posthumous" "Nobel Prize" "Nobel Prize" "Nobel Prize" ...
##  $ Education         : chr  "Self-taught" "Ph.D. in Astronomy" "Ph.D. in Mathematics" "University of
##  $ Influence         : chr  "Popularizing science and cosmology" "Foundational work in quantum mechan
##  $ NobelPrize        : num  0 1 1 1 1 1 0 1 0 0 ...
```

```r
table(topIntelligent_cleaned$NobelPrize)
```

```
##
##    0    1
## 1800 1951
```

```r
table(topIntelligent_cleaned$NobelPrize, topIntelligent_cleaned$Country)
```

```
##
##     Austria Denmark France Germany Italy  UK USA
##   0      92      86     91      93   273 763 402
##   1      96     106     92     104   320 872 361
```

I need to convert categorical variables from string format to factors for the logistic regression model.

```r
topIntelligent_cleaned$Country <- as.factor(topIntelligent_cleaned$Country)
topIntelligent_cleaned$Field.of.Expertise <- as.factor(topIntelligent_cleaned$Field.of.Expertise)
topIntelligent_cleaned$Achievements <- as.factor(topIntelligent_cleaned$Achievements)
topIntelligent_cleaned$Gender <- as.factor(topIntelligent_cleaned$Gender)
topIntelligent_cleaned$Education <- as.factor(topIntelligent_cleaned$Education)
topIntelligent_cleaned$Influence <- as.factor(topIntelligent_cleaned$Influence)
```

Evaluate Model Fit Using Backward Selection I will evaluate the model fit using a backward selection method. I include all variables in the model and remove them one at a time, checking the model fit using AIC (Akaike Information Criterion) and Deviance.

```r
backward_model <- glm(NobelPrize ~ Country + Field.of.Expertise + Achievements + Gender + Education + In

stepwise_backward_model <- step(backward_model, direction = "backward")
```

```
## Start:  AIC=5252.2
## NobelPrize ~ Country + Field.of.Expertise + Achievements + Gender +
##     Education + Influence + IQ
##
##                      Df Deviance    AIC
## - Influence          18   5149.1 5231.1
## - Achievements       18   5152.7 5234.7
## - Education           8   5139.4 5241.4
## - Field.of.Expertise  6   5139.1 5245.1
## - Gender              1   5134.3 5250.3
## - Country             6   5144.6 5250.6
## <none>                    5134.2 5252.2
## - IQ                  1   5138.4 5254.4
```

```
##
## Step:  AIC=5231.07
## NobelPrize ~ Country + Field.of.Expertise + Achievements + Gender +
##     Education + IQ
##
##                      Df Deviance    AIC
## - Achievements       18   5168.2 5214.2
## - Education           8   5154.7 5220.7
## - Field.of.Expertise  6   5153.8 5223.8
## - Country             6   5158.9 5228.9
## - Gender              1   5149.2 5229.2
## <none>                    5149.1 5231.1
## - IQ                  1   5153.5 5233.5
##
## Step:  AIC=5214.17
## NobelPrize ~ Country + Field.of.Expertise + Gender + Education +
##     IQ
##
##                      Df Deviance    AIC
## - Education           8   5174.2 5204.2
## - Field.of.Expertise  6   5172.7 5206.7
## - Country             6   5178.0 5212.0
## - Gender              1   5168.3 5212.3
## <none>                    5168.2 5214.2
## - IQ                  1   5173.1 5217.1
##
## Step:  AIC=5204.23
## NobelPrize ~ Country + Field.of.Expertise + Gender + IQ
##
##                      Df Deviance    AIC
## - Field.of.Expertise  6   5178.9 5196.9
## - Country             6   5183.9 5201.9
## - Gender              1   5174.3 5202.3
## <none>                    5174.2 5204.2
## - IQ                  1   5179.1 5207.1
##
## Step:  AIC=5196.92
## NobelPrize ~ Country + Gender + IQ
##
##           Df Deviance    AIC
## - Country  6   5188.6 5194.6
## - Gender   1   5179.0 5195.0
## <none>         5178.9 5196.9
## - IQ       1   5183.9 5199.9
##
## Step:  AIC=5194.58
## NobelPrize ~ Gender + IQ
##
##          Df Deviance    AIC
## - Gender  1   5188.6 5192.6
## <none>        5188.6 5194.6
## - IQ      1   5193.9 5197.9
##
## Step:  AIC=5192.61
```

```
## NobelPrize ~ IQ
##
##        Df Deviance    AIC
## <none>      5188.6 5192.6
## - IQ    1   5193.9 5195.9
```

```r
# summary(backward_model)
```

Interpretation: Only IQ turns out to be a significant in predicting to Nobel Prize win. Removing this variable does degrade the model fit (i.e., does increase AIC).

- Validating significant predictor

```r
# Confidence intervals for the model coefficients
confint.default(backward_model)
```

```
##                                                              2.5 %      97.5 %
## (Intercept)                                          -1.2814388806 0.316324656
## CountryDenmark                                       -0.2168975272 0.600379277
## CountryFrance                                        -0.4072748844 0.417343014
## CountryGermany                                       -0.3122882607 0.497785336
## CountryItaly                                         -0.1943015188 0.472964017
## CountryUK                                            -0.1896435480 0.423346397
## CountryUSA                                           -0.4599122747 0.187571220
## Field.of.ExpertiseBiology                            -0.3438846059 0.355746205
## Field.of.ExpertiseChemistry                          -0.2449495012 0.318818286
## Field.of.ExpertiseEngineering                        -0.5484936715 0.142514144
## Field.of.ExpertiseMathematics                        -0.3450502443 0.150329188
## Field.of.ExpertisePhysics                            -0.1789892788 0.262519740
## Field.of.ExpertisePolymath                           -0.2622145707 0.449988003
## AchievementsBlack Hole Theory                        -0.6356992654 0.182882957
## AchievementsCosmos Series                            -0.2817011738 0.551264172
## AchievementsDiscovery of Radium and Polonium         -0.6960395716 0.123179793
## AchievementsElectromagnetic Induction                -0.4365755055 0.393167295
## AchievementsElectromagnetic Theory                   -0.6991410447 0.095210153
## AchievementsFather of Computer Science               -0.3584179345 0.433916540
## AchievementsFirst computer algorithm                 -0.4740784939 0.356792664
## AchievementsGeneral Theory of Relativity             -0.5124548012 0.316615610
## AchievementsHeliocentric Theory                      -0.5799011980 0.244173246
## AchievementsLaws of Motion                           -0.4517383545 0.388297275
## AchievementsMona Lisa, The Last Supper, Inventions   -0.5250938284 0.334037978
## AchievementsNuclear Reactor                          -0.6423370035 0.188954580
## AchievementsQuantum Electrodynamics                  -0.6496881600 0.157780800
## AchievementsQuantum Mechanics                        -0.3725236876 0.333043625
## AchievementsQuantum Theory                           -0.6850974970 0.116966154
## AchievementsTheory of Evolution                      -0.7357649227 0.061201686
## AchievementsTheory of Relativity                     -0.7275852817 0.102274733
## AchievementsWave Equation                            -0.3405349110 0.479743978
## GenderMale                                           -0.1086925175 0.151522783
## EducationPh.D. (honorary)                            -0.3991021956 0.437354638
## EducationPh.D. in Astronomy                          -0.5473267586 0.291451391
## EducationPh.D. in Chemistry                          -0.3946227460 0.344352975
## EducationPh.D. in Mathematics                        -0.2973266637 0.408459072
## EducationPh.D. in Physics                            -0.3959756116 0.258538717
## EducationSelf-taught                                 -0.3069932148 0.398599390
## EducationUniversity of Cambridge                     -0.2562019828 0.598776901
```

```
## EducationUniversity of Pisa                          -0.5939163165 0.274446419
## InfluenceDeveloped quantum theory                    -0.2731433240 0.547843626
## InfluenceDiscovery of electromagnetic induction      -0.4311737959 0.373431393
## InfluenceEarly pioneer of computer science           -0.4404523598 0.371793108
## InfluenceEvolutionary biology pioneer                -0.2828577277 0.512223206
## InfluenceFoundation of classical mechanics           -0.1287462939 0.700591164
## InfluenceFoundational work in quantum mechanics      -0.3911967362 0.317735152
## InfluenceIconic Renaissance artist and inventor      -0.3918493985 0.404112724
## InfluenceInventor and electrical engineering pioneer -0.1199211015 0.708563506
## InfluenceLaid the groundwork for modern computing    -0.4094507665 0.385622285
## InfluenceMajor advancements in astronomy             -0.1403882635 0.686673743
## InfluenceNuclear physics and reactor development     -0.2709516889 0.541641642
## InfluencePenrose's work on black holes and cosmology -0.1682656425 0.660940447
## InfluencePioneering research in radioactivity        -0.2162453346 0.616659132
## InfluencePopularizing science and cosmology          -0.2401902149 0.567849763
## InfluenceQuantum electrodynamics and physics teaching -0.3249113308 0.477554450
## InfluenceRevolutionized modern physics               -0.5448577665 0.253519052
## InfluenceUnified theory of electromagnetism          -0.3999363727 0.428714497
## InfluenceWave mechanics in quantum physics           -0.1737790942 0.642286760
## IQ                                                    0.0001412904 0.006502679
```

```
# Exponentiated coefficients
exp(coef(backward_model))
```

```
##                                             (Intercept)
##                                               0.6172031
##                                         CountryDenmark
##                                               1.2113566
##                                          CountryFrance
##                                               1.0050468
##                                         CountryGermany
##                                               1.0971858
##                                           CountryItaly
##                                               1.1495048
##                                              CountryUK
##                                               1.1239524
##                                             CountryUSA
##                                               0.8726938
##                               Field.of.ExpertiseBiology
##                                               1.0059484
##                             Field.of.ExpertiseChemistry
##                                               1.0376249
##                           Field.of.ExpertiseEngineering
##                                               0.8162866
##                           Field.of.ExpertiseMathematics
##                                               0.9072289
##                               Field.of.ExpertisePhysics
##                                               1.0426497
##                              Field.of.ExpertisePolymath
##                                               1.0984353
##                           AchievementsBlack Hole Theory
##                                               0.7973926
##                               AchievementsCosmos Series
##                                               1.1442867
##            AchievementsDiscovery of Radium and Polonium
```

```
##                                            0.7509397
##                 AchievementsElectromagnetic Induction
##                                            0.9785297
##                    AchievementsElectromagnetic Theory
##                                            0.7393636
##                AchievementsFather of Computer Science
##                                            1.0384709
##                  AchievementsFirst computer algorithm
##                                            0.9430435
##              AchievementsGeneral Theory of Relativity
##                                            0.9067218
##                       AchievementsHeliocentric Theory
##                                            0.8454688
##                            AchievementsLaws of Motion
##                                            0.9687773
##    AchievementsMona Lisa, The Last Supper, Inventions
##                                            0.9088930
##                           AchievementsNuclear Reactor
##                                            0.7971669
##                   AchievementsQuantum Electrodynamics
##                                            0.7819584
##                         AchievementsQuantum Mechanics
##                                            0.9804535
##                            AchievementsQuantum Theory
##                                            0.7527172
##                       AchievementsTheory of Evolution
##                                            0.7137078
##                      AchievementsTheory of Relativity
##                                            0.7315020
##                             AchievementsWave Equation
##                                            1.0720841
##                                           GenderMale
##                                            1.0216461
##                             EducationPh.D. (honorary)
##                                            1.0193103
##                           EducationPh.D. in Astronomy
##                                            0.8799082
##                           EducationPh.D. in Chemistry
##                                            0.9751784
##                         EducationPh.D. in Mathematics
##                                            1.0571390
##                             EducationPh.D. in Physics
##                                            0.9335895
##                                  EducationSelf-taught
##                                            1.0468682
##                      EducationUniversity of Cambridge
##                                            1.1868319
##                           EducationUniversity of Pisa
##                                            0.8523697
##                     InfluenceDeveloped quantum theory
##                                            1.1472298
##       InfluenceDiscovery of electromagnetic induction
##                                            0.9715416
##            InfluenceEarly pioneer of computer science
```

```
##                                                     0.9662530
##                  InfluenceEvolutionary biology pioneer
##                                                     1.1215176
##             InfluenceFoundation of classical mechanics
##                                                     1.3309892
##        InfluenceFoundational work in quantum mechanics
##                                                     0.9639356
##        InfluenceIconic Renaissance artist and inventor
##                                                     1.0061505
##   InfluenceInventor and electrical engineering pioneer
##                                                     1.3422150
##      InfluenceLaid the groundwork for modern computing
##                                                     0.9881565
##               InfluenceMajor advancements in astronomy
##                                                     1.3140878
##       InfluenceNuclear physics and reactor development
##                                                     1.1449317
##   InfluencePenrose's work on black holes and cosmology
##                                                     1.2793311
##          InfluencePioneering research in radioactivity
##                                                     1.2216555
##            InfluencePopularizing science and cosmology
##                                                     1.1780138
## InfluenceQuantum electrodynamics and physics teaching
##                                                     1.0793096
##                 InfluenceRevolutionized modern physics
##                                                     0.8644435
##            InfluenceUnified theory of electromagnetism
##                                                     1.0144931
##             InfluenceWave mechanics in quantum physics
##                                                     1.2639653
##                                                            IQ
##                                                     1.0033275
```

```r
library(broom)
tidy(stepwise_backward_model, exponentiate = TRUE)
```

```
## # A tibble: 2 x 5
##   term        estimate std.error statistic p.value
##   <chr>          <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept)    0.591   0.266       -1.98  0.0477
## 2 IQ             1.00    0.00160      2.30  0.0214
```

- To validate regression model output, now I will use Chi-Squared test to find association between the outcome and some categorical variables at a time.

```r
chisq.test(table(topIntelligent_cleaned$Country, topIntelligent_cleaned$NobelPrize))
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(topIntelligent_cleaned$Country, topIntelligent_cleaned$NobelPrize)
## X-squared = 9.9205, df = 6, p-value = 0.128
```

```r
chisq.test(table(topIntelligent_cleaned$Gender, topIntelligent_cleaned$NobelPrize))
```

```
##
```

```
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(topIntelligent_cleaned$Gender, topIntelligent_cleaned$NobelPrize)
## X-squared = 0.017409, df = 1, p-value = 0.895
```

```r
chisq.test(table(topIntelligent_cleaned$Education, topIntelligent_cleaned$NobelPrize))
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(topIntelligent_cleaned$Education, topIntelligent_cleaned$NobelPrize)
## X-squared = 6.0064, df = 8, p-value = 0.6465
```

```r
chisq.test(table(topIntelligent_cleaned$Field.of.Expertise, topIntelligent_cleaned$NobelPrize))
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(topIntelligent_cleaned$Field.of.Expertise, topIntelligent_cleaned$NobelPrize)
## X-squared = 4.7284, df = 6, p-value = 0.5791
```

```r
chisq.test(table(topIntelligent_cleaned$Achievements, topIntelligent_cleaned$NobelPrize))
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(topIntelligent_cleaned$Achievements, topIntelligent_cleaned$NobelPrize)
## X-squared = 19.903, df = 18, p-value = 0.3383
```

Interpretation: Again, no categorical variables under investigation are significant in predicting getting a Nobel Prize.

4.1 Lasso Regression Analysis

I will try out Lasso Regression to find which variables are effective in predicting a Nobel Prize win.

```r
# install.packages("glmnet")
```

```r
library(glmnet)
```

```
## Loading required package: Matrix

##
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack

## Loaded glmnet 4.1-8
```

```r
data_dummy <- model.matrix(NobelPrize ~ Country + Field.of.Expertise + Achievements + Gender + Educatio
```

```r
X <- data_dummy
y <- topIntelligent_cleaned$NobelPrize
```

```r
lasso_model <- cv.glmnet(X, y, family = "binomial", alpha = 1)
```

```r
best_lambda <- lasso_model$lambda.min
coef(lasso_model, s = best_lambda)
```

```
## 60 x 1 sparse Matrix of class "dgCMatrix"
##                                                              s1
## (Intercept)                                          0.08482707
## CountryAustria                                       .
## CountryDenmark                                       .
## CountryFrance                                        .
## CountryGermany                                       .
## CountryItaly                                         .
## CountryUK                                            .
## CountryUSA                                          -0.02099338
## Field.of.ExpertiseBiology                            .
## Field.of.ExpertiseChemistry                          .
## Field.of.ExpertiseEngineering                        .
## Field.of.ExpertiseMathematics                        .
## Field.of.ExpertisePhysics                            .
## Field.of.ExpertisePolymath                           .
## AchievementsBlack Hole Theory                        .
## AchievementsCosmos Series                            .
## AchievementsDiscovery of Radium and Polonium         .
## AchievementsElectromagnetic Induction                .
## AchievementsElectromagnetic Theory                   .
## AchievementsFather of Computer Science               .
## AchievementsFirst computer algorithm                 .
## AchievementsGeneral Theory of Relativity             .
## AchievementsHeliocentric Theory                      .
## AchievementsLaws of Motion                           .
## AchievementsMona Lisa, The Last Supper, Inventions   .
## AchievementsNuclear Reactor                          .
## AchievementsQuantum Electrodynamics                  .
## AchievementsQuantum Mechanics                        .
## AchievementsQuantum Theory                           .
## AchievementsTheory of Evolution                      .
## AchievementsTheory of Relativity                     .
## AchievementsWave Equation                            .
## GenderMale                                           .
## EducationPh.D. (honorary)                            .
## EducationPh.D. in Astronomy                          .
## EducationPh.D. in Chemistry                          .
## EducationPh.D. in Mathematics                        .
## EducationPh.D. in Physics                            .
## EducationSelf-taught                                 .
## EducationUniversity of Cambridge                     .
## EducationUniversity of Pisa                          .
## InfluenceDeveloped quantum theory                    .
## InfluenceDiscovery of electromagnetic induction      .
## InfluenceEarly pioneer of computer science           .
## InfluenceEvolutionary biology pioneer                .
## InfluenceFoundation of classical mechanics           .
## InfluenceFoundational work in quantum mechanics      .
## InfluenceIconic Renaissance artist and inventor      .
## InfluenceInventor and electrical engineering pioneer .
## InfluenceLaid the groundwork for modern computing    .
## InfluenceMajor advancements in astronomy             .
## InfluenceNuclear physics and reactor development     .
```

```
## InfluencePenrose's work on black holes and cosmology   .
## InfluencePioneering research in radioactivity          .
## InfluencePopularizing science and cosmology            .
## InfluenceQuantum electrodynamics and physics teaching  .
## InfluenceRevolutionized modern physics                 .
## InfluenceUnified theory of electromagnetism            .
## InfluenceWave mechanics in quantum physics             .
## IQ                                                     .
```

Interpretation:

Lasso model finds the most relevant predictors to the outcome variable. Here, non-zero coefficients contribute predicting Nobel Prize win. Only the country USA and IQ have non-zero coefficients. And, IQ slightly affects the Nobel prize win: Higher IQ is associated with a very slight increase in the log-odds of receiving a Nobel Prize. On the other hand, being a US citizen decreases receiving a Nobel Prize, which can be confirmed by the following contingency table.

```
table(topIntelligent_cleaned$Country, topIntelligent_cleaned$NobelPrize)
```

```
##
##             0   1
##   Austria  92  96
##   Denmark  86 106
##   France   91  92
##   Germany  93 104
##   Italy   273 320
##   UK      763 872
##   USA     402 361
```

Summary:

In this project, I have worked on Top Intelligent People in the world and and tried to gain some insight on their profile using EDA and regression analysis. EDA has provided descriptive information on questions regarding sinlge and multi-variables. Some of them include Distribution of Awards by Country, Distribution of Awards by Country, IQ and its relation with some variables (e.g., Awards, Education, Country etc.), Distribution of Awards by Fields. Questions in EDA can be expanded depending on the needs.

In the last section, I was curious about what predictors may contribute winning a Nobel Prize. The results from Logistic Regression showed that only IQ is significant in predicting winning a Nobel Prize.

Further, I wanted to validate the results of Logistic Regression by building a Lasso Regression model, which helps identifying which predictors (variables) are most important by shrinking the less significant coefficients to zero. This model is considered to be useful in datasets with many variables or multicollinear variables, as it can help simplify the model by focusing on the most relevant features/variables. Simply, with Lasso Regression, I wanted to increase the predictive accuracy of my model by selecting important variables, and to better understand the key factors influencing Nobel Prize achievements.

The results of the Lasso Regression model revealed that IQ has slightly influential in predicting Nobel Prize win, while the country USA has negatively associated with winning a Nobel Prize.