Nobel_Prizes_Analysis

November 19, 2025

1 Setup and Context

1.0.1 Introduction

On November 27, 1895, Alfred Nobel signed his last will in Paris. When it was opened after his death, the will caused a lot of controversy, as Nobel had left much of his wealth for the establishment of a prize.

Alfred Nobel dictates that his entire remaining estate should be used to endow "prizes to those who, during the preceding year, have conferred the greatest benefit to humankind".

Every year the Nobel Prize is given to scientists and scholars in the categories chemistry, literature, physics, physiology or medicine, economics, and peace.

Let's see what patterns we can find in the data of the past Nobel laureates. What can we learn about the Nobel prize and our world more generally?

1.0.2 Import Statements

[1]:
```
import pandas as pd
import numpy as np
import plotly.express as px
import seaborn as sns
import matplotlib.pyplot as plt
```

1.0.3 Notebook Presentation

[2]:
```
pd.options.display.float_format = '{:,.2f}'.format
```

1.0.4 Read the Data

[3]:
```
df_data = pd.read_csv('nobel_prize_data.csv')
```

1.0.5 Explore the DataFrame

[4]:
```
df_data.shape
```

[4]:
```
(962, 16)
```

[5]:
```
df_data.columns
```

[5]:
```
Index(['year', 'category', 'prize', 'motivation', 'prize_share',
       'laureate_type', 'full_name', 'birth_date', 'birth_city',
       'birth_country', 'birth_country_current', 'sex', 'organization_name',
       'organization_city', 'organization_country', 'ISO'],
      dtype='object')
```

[6]:
```
df_data.head()
```

[6]:

	year	category	prize	motivation	prize_share	laureate_type	full_name	birth_date	birth_city	birth_country	birth_country_current	sex	organization_name	organization_city	organization_country	ISO
0	1901	Chemistry	The Nobel Prize in Chemistry 1901	"in recognition of the extraordinary services …	1/1	Individual	Jacobus Henricus van 't Hoff	1852-08-30	Rotterdam	Netherlands	Netherlands	Male	Berlin University	Berlin	Germany	NLD
1	1901	Literature	The Nobel Prize in Literature 1901	"in special recognition of his poetic composit…	1/1	Individual	Sully Prudhomme	1839-03-16	Paris	France	France	Male	NaN	NaN	NaN	FRA
2	1901	Medicine	The Nobel Prize in Physiology or Medicine 1901	"for his work on serum therapy, especially its…	1/1	Individual	Emil Adolf von Behring	1854-03-15	Hansdorf (Lawice)	Prussia (Poland)	Poland	Male	Marburg University	Marburg	Germany	POL
3	1901	Peace	The Nobel Peace Prize 1901	NaN	1/2	Individual	Frédéric Passy	1822-05-20	Paris	France	France	Male	NaN	NaN	NaN	FRA
4	1901	Peace	The Nobel Peace Prize 1901	NaN	1/2	Individual	Jean Henry Dunant	1828-05-08	Geneva	Switzerland	Switzerland	Male	NaN	NaN	NaN	CHE

[7]:
```
df_data.tail()
```

[7]:

	year	category	prize	motivation	prize_share	laureate_type	full_name	birth_date	birth_city	birth_country	birth_country_current	sex	organization_name	organization_city	organization_country	ISO
957	2020	Medicine	The Nobel Prize in Physiology or Medicine 2020													

```
958  2020     Peace                      The Nobel Peace Prize 2020
959  2020   Physics                  The Nobel Prize in Physics 2020
960  2020   Physics                  The Nobel Prize in Physics 2020
961  2020   Physics                  The Nobel Prize in Physics 2020

                                          motivation prize_share  \
957          "for the discovery of Hepatitis C virus"         1/3
958  "for its efforts to combat hunger, for its con…         1/1
959  "for the discovery of a supermassive compact o…         1/4
960  "for the discovery of a supermassive compact o…         1/4
961  "for the discovery that black hole formation i…         1/2

    laureate_type                   full_name  birth_date  \
957    Individual            Michael Houghton  1949-07-02
958  Organization  World Food Programme (WFP)         NaN
959    Individual                 Andrea Ghez  1965-06-16
960    Individual             Reinhard Genzel  1952-03-24
961    Individual               Roger Penrose  1931-08-08

                  birth_city             birth_country  \
957                      NaN            United Kingdom
958                      NaN                       NaN
959             New York, NY  United States of America
960  Bad Homburg vor der Höhe                   Germany
961               Colchester            United Kingdom

       birth_country_current     sex         organization_name  \
957            United Kingdom    Male     University of Alberta
958                       NaN     NaN                       NaN
959  United States of America  Female  University of California
960                   Germany    Male  University of California
961            United Kingdom    Male      University of Oxford

    organization_city      organization_country  ISO
957          Edmonton                    Canada  GBR
958               NaN                       NaN  NaN
959      Berkeley, CA  United States of America  USA
960   Los Angeles, CA  United States of America  DEU
961            Oxford            United Kingdom  GBR
```

1.0.6 Check for Duplicates

[8]:
```
df_data.duplicated().values.any()
```

[8]:
```
False
```

[9]:
```
print(f"There are duplicates : {df_data.duplicated().values.any()}")
```

```
There are duplicates : False
```

1.0.7 Check for NaN Values

[10]:
```
print(f"There are NaN values : {df_data.isna().values.any()}")
```

```
There are NaN values : True
```

[11]:
```
df_data.isna().sum()
```

[11]:
```
year                     0
category                 0
prize                    0
motivation              88
prize_share              0
laureate_type            0
full_name                0
birth_date              28
birth_city              31
birth_country           28
birth_country_current   28
sex                     28
organization_name      255
organization_city      255
organization_country   254
ISO                     28
dtype: int64
```

[12]:
```
col_subset = ['year', 'category', 'laureate_type', 'birth_date', 'full_name',␣
↪'organization_name']
```

[13]:
```
df_data.loc[df_data.birth_date.isna()][col_subset]
```

[13]:

	year	category	laureate_type	birth_date	\
24	1904	Peace	Organization	NaN	
60	1910	Peace	Organization	NaN	
89	1917	Peace	Organization	NaN	
200	1938	Peace	Organization	NaN	
215	1944	Peace	Organization	NaN	
237	1947	Peace	Organization	NaN	
238	1947	Peace	Organization	NaN	
283	1954	Peace	Organization	NaN	
348	1963	Peace	Organization	NaN	
349	1963	Peace	Organization	NaN	
366	1965	Peace	Organization	NaN	
399	1969	Peace	Organization	NaN	
479	1977	Peace	Organization	NaN	
523	1981	Peace	Organization	NaN	

```
558  1985    Peace  Organization        NaN
588  1988    Peace  Organization        NaN
659  1995    Peace  Organization        NaN
682  1997    Peace  Organization        NaN
703  1999    Peace  Organization        NaN
730  2001    Peace  Organization        NaN
778  2005    Peace  Organization        NaN
788  2006    Peace  Organization        NaN
801  2007    Peace  Organization        NaN
860  2012    Peace  Organization        NaN
873  2013    Peace  Organization        NaN
897  2015    Peace  Organization        NaN
919  2017    Peace  Organization        NaN
958  2020    Peace  Organization        NaN
```

	full_name	organization_name
24	Institut de droit international (Institute of …	NaN
60	Bureau international permanent de la Paix (Per…	NaN
89	Comité international de la Croix Rouge (Intern…	NaN
200	Office international Nansen pour les Réfugiés …	NaN
215	Comité international de la Croix Rouge (Intern…	NaN
237	American Friends Service Committee (The Quakers)	NaN
238	Friends Service Council (The Quakers)	NaN
283	Office of the United Nations High Commissioner…	NaN
348	Comité international de la Croix Rouge (Intern…	NaN
349	Ligue des Sociétés de la Croix-Rouge (League o…	NaN
366	United Nations Children's Fund (UNICEF)	NaN
399	International Labour Organization (I.L.O.)	NaN
479	Amnesty International	NaN
523	Office of the United Nations High Commissioner…	NaN
558	International Physicians for the Prevention of…	NaN
588	United Nations Peacekeeping Forces	NaN
659	Pugwash Conferences on Science and World Affairs	NaN
682	International Campaign to Ban Landmines (ICBL)	NaN
703	Médecins Sans Frontières	NaN
730	United Nations (U.N.)	NaN
778	International Atomic Energy Agency (IAEA)	NaN
788	Grameen Bank	NaN
801	Intergovernmental Panel on Climate Change (IPCC)	NaN
860	European Union (EU)	NaN
873	Organisation for the Prohibition of Chemical W…	NaN
897	National Dialogue Quartet	NaN
919	International Campaign to Abolish Nuclear Weap…	NaN
958	World Food Programme (WFP)	NaN

[14]:
```
df_data.loc[df_data.organization_name.isna()][col_subset]
```

[14]:
```
      year    category laureate_type  birth_date  \
1     1901  Literature    Individual  1839-03-16
3     1901       Peace    Individual  1822-05-20
4     1901       Peace    Individual  1828-05-08
7     1902  Literature    Individual  1817-11-30
9     1902       Peace    Individual  1843-05-21
..     …          …            …          …
932   2018       Peace    Individual  1993-07-02
942   2019  Literature    Individual  1942-12-06
946   2019       Peace    Individual  1976-08-15
954   2020  Literature    Individual  1943-04-22
958   2020       Peace  Organization         NaN

                              full_name organization_name
1                        Sully Prudhomme               NaN
3                         Frédéric Passy               NaN
4                      Jean Henry Dunant               NaN
7    Christian Matthias Theodor Mommsen               NaN
9                    Charles Albert Gobat               NaN
..                                    …               …
932                         Nadia Murad               NaN
942                        Peter Handke               NaN
946                       Abiy Ahmed Ali               NaN
954                         Louise Glück               NaN
958          World Food Programme (WFP)               NaN

[255 rows x 6 columns]
```

Convert Year and Birth Date to Datetime

[15]:
```
df_data.birth_date = pd.to_datetime(df_data.birth_date)
```

[16]:
```
df_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 962 entries, 0 to 961
Data columns (total 16 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   year                  962 non-null    int64
 1   category              962 non-null    object
 2   prize                 962 non-null    object
 3   motivation            874 non-null    object
 4   prize_share           962 non-null    object
 5   laureate_type         962 non-null    object
 6   full_name             962 non-null    object
 7   birth_date            934 non-null    datetime64[ns]
 8   birth_city            931 non-null    object
 9   birth_country         934 non-null    object
```

```
 10  birth_country_current  934 non-null    object
 11  sex                    934 non-null    object
 12  organization_name      707 non-null    object
 13  organization_city      707 non-null    object
 14  organization_country   708 non-null    object
 15  ISO                    934 non-null    object
dtypes: datetime64[ns](1), int64(1), object(14)
memory usage: 120.4+ KB
```

Add a Column with the Prize Share as a Percentage

[17]:
```
separated_values = df_data.prize_share.str.split('/', expand = True)
numerator = pd.to_numeric(separated_values[0])
denomerator = pd.to_numeric(separated_values[1])
df_data['share_pct'] = numerator / denomerator
```

2 Plotly Donut Chart: Percentage of Male vs. Female Laureates

[19]:
```
sex = df_data.sex.value_counts()
```

[20]:
```
sex
```

[20]:
```
sex
Male      876
Female     58
Name: count, dtype: int64
```

[21]:
```
fig = px.pie(labels=sex.index,
             values=sex.values,
             title="Percentage of Male vs. Female Winners",
             names=sex.index,
             hole=0.4,)

fig.update_traces(textposition='inside', textfont_size=15, textinfo='percent')

fig.show()
```

3 Who were the first 3 Women to Win the Nobel Prize?

[32]:
```
df_data
```

[32]:

	year	category	prize	\
0	1901	Chemistry	The Nobel Prize in Chemistry 1901	
1	1901	Literature	The Nobel Prize in Literature 1901	
2	1901	Medicine	The Nobel Prize in Physiology or Medicine 1901	
3	1901	Peace	The Nobel Peace Prize 1901	
4	1901	Peace	The Nobel Peace Prize 1901	

```
..    …         …                                                 …
957  2020    Medicine  The Nobel Prize in Physiology or Medicine 2020
958  2020       Peace                      The Nobel Peace Prize 2020
959  2020     Physics                The Nobel Prize in Physics 2020
960  2020     Physics                The Nobel Prize in Physics 2020
961  2020     Physics                The Nobel Prize in Physics 2020

                                          motivation prize_share  \
0    "in recognition of the extraordinary services …         1/1
1    "in special recognition of his poetic composit…         1/1
2    "for his work on serum therapy, especially its…         1/1
3                                                NaN         1/2
4                                                NaN         1/2
..                                               …         …
957          "for the discovery of Hepatitis C virus"         1/3
958  "for its efforts to combat hunger, for its con…         1/1
959  "for the discovery of a supermassive compact o…         1/4
960  "for the discovery of a supermassive compact o…         1/4
961  "for the discovery that black hole formation i…         1/2

    laureate_type                    full_name birth_date  \
0      Individual  Jacobus Henricus van 't Hoff 1852-08-30
1      Individual               Sully Prudhomme 1839-03-16
2      Individual        Emil Adolf von Behring 1854-03-15
3      Individual                Frédéric Passy 1822-05-20
4      Individual             Jean Henry Dunant 1828-05-08
..            …                           …       …
957    Individual              Michael Houghton 1949-07-02
958  Organization    World Food Programme (WFP)        NaT
959    Individual                  Andrea Ghez 1965-06-16
960    Individual               Reinhard Genzel 1952-03-24
961    Individual                 Roger Penrose 1931-08-08

                  birth_city             birth_country  \
0                  Rotterdam               Netherlands
1                      Paris                    France
2          Hansdorf (Lawice)          Prussia (Poland)
3                      Paris                    France
4                     Geneva               Switzerland
..                       …                       …
957                     NaN            United Kingdom
958                     NaN                       NaN
959             New York, NY  United States of America
960  Bad Homburg vor der Höhe                 Germany
961               Colchester            United Kingdom

        birth_country_current     sex         organization_name  \
```

```
0                  Netherlands    Male         Berlin University
1                       France    Male                        NaN
2                       Poland    Male       Marburg University
3                       France    Male                        NaN
4                  Switzerland    Male                        NaN
..                         …     …                          …
957              United Kingdom    Male     University of Alberta
958                         NaN     NaN                        NaN
959  United States of America  Female  University of California
960                     Germany    Male  University of California
961              United Kingdom    Male      University of Oxford

    organization_city      organization_country  ISO  share_pct
0              Berlin                   Germany  NLD       1.00
1                 NaN                       NaN  FRA       1.00
2             Marburg                   Germany  POL       1.00
3                 NaN                       NaN  FRA       0.50
4                 NaN                       NaN  CHE       0.50
..                …                         …  …        …
957           Edmonton                  Canada  GBR       0.33
958               NaN                       NaN  NaN       1.00
959      Berkeley, CA  United States of America  USA       0.25
960   Los Angeles, CA  United States of America  DEU       0.25
961            Oxford            United Kingdom  GBR       0.50

[962 rows x 17 columns]
```

[33]:
```
df_women.sort_values('year', ascending = True).head(3)
```

[33]:
```
    year    category                                prize  \
18  1903     Physics     The Nobel Prize in Physics 1903
29  1905       Peace          The Nobel Peace Prize 1905
51  1909  Literature  The Nobel Prize in Literature 1909

                                          motivation prize_share  \
18  "in recognition of the extraordinary services …         1/4
29                                              NaN         1/1
51  "in appreciation of the lofty idealism, vivid …         1/1

   laureate_type                                          full_name  \
18    Individual                       Marie Curie, née Sklodowska
29    Individual  Baroness Bertha Sophie Felicita von Suttner, n…
51    Individual                      Selma Ottilia Lovisa Lagerlöf

   birth_date birth_city                    birth_country  \
18 1867-11-07     Warsaw           Russian Empire (Poland)
29 1843-06-09     Prague  Austrian Empire (Czech Republic)
```

```
51 1858-11-20   Mårbacka                                  Sweden

   birth_country_current     sex organization_name organization_city  \
18                Poland  Female               NaN               NaN
29        Czech Republic  Female               NaN               NaN
51                Sweden  Female               NaN               NaN

   organization_country  ISO  share_pct
18                   NaN  POL       0.25
29                   NaN  CZE       1.00
51                   NaN  SWE       1.00
```

4 Repeat Winners

[26]:
```
is_winner = df_data.duplicated(subset=['full_name'], keep=False)
```

[27]:
```
multiple_winners = df_data[is_winner]
```

[28]:
```
multiple_winners
```

[28]:
```
     year   category                              prize  \
18   1903    Physics    The Nobel Prize in Physics 1903
62   1911  Chemistry  The Nobel Prize in Chemistry 1911
89   1917      Peace         The Nobel Peace Prize 1917
215  1944      Peace         The Nobel Peace Prize 1944
278  1954  Chemistry  The Nobel Prize in Chemistry 1954
283  1954      Peace         The Nobel Peace Prize 1954
297  1956    Physics    The Nobel Prize in Physics 1956
306  1958  Chemistry  The Nobel Prize in Chemistry 1958
340  1962      Peace         The Nobel Peace Prize 1962
348  1963      Peace         The Nobel Peace Prize 1963
424  1972    Physics    The Nobel Prize in Physics 1972
505  1980  Chemistry  The Nobel Prize in Chemistry 1980
523  1981      Peace         The Nobel Peace Prize 1981

                                          motivation prize_share  \
18   "in recognition of the extraordinary services …         1/4
62   "in recognition of her services to the advance…         1/1
89                                               NaN         1/1
215                                              NaN         1/1
278  "for his research into the nature of the chemi…         1/1
283                                              NaN         1/1
297  "for their researches on semiconductors and th…         1/3
306  "for his work on the structure of proteins, es…         1/1
340                                              NaN         1/1
348                                              NaN         1/2
```

```
424  "for their jointly developed theory of superco…         1/3
505  "for their contributions concerning the determ…         1/4
523                                                NaN         1/1

    laureate_type                                          full_name  \
18     Individual                        Marie Curie, née Sklodowska
62     Individual                        Marie Curie, née Sklodowska
89   Organization  Comité international de la Croix Rouge (Intern…
215  Organization  Comité international de la Croix Rouge (Intern…
278    Individual                                  Linus Carl Pauling
283  Organization  Office of the United Nations High Commissioner…
297    Individual                                         John Bardeen
306    Individual                                   Frederick Sanger
340    Individual                                  Linus Carl Pauling
348  Organization  Comité international de la Croix Rouge (Intern…
424    Individual                                         John Bardeen
505    Individual                                   Frederick Sanger
523  Organization  Office of the United Nations High Commissioner…

    birth_date    birth_city             birth_country  \
18  1867-11-07        Warsaw   Russian Empire (Poland)
62  1867-11-07        Warsaw   Russian Empire (Poland)
89         NaT           NaN                       NaN
215        NaT           NaN                       NaN
278 1901-02-28  Portland, OR  United States of America
283        NaT           NaN                       NaN
297 1908-05-23   Madison, WI  United States of America
306 1918-08-13     Rendcombe            United Kingdom
340 1901-02-28  Portland, OR  United States of America
348        NaT           NaN                       NaN
424 1908-05-23   Madison, WI  United States of America
505 1918-08-13     Rendcombe            United Kingdom
523        NaT           NaN                       NaN

        birth_country_current     sex  \
18                     Poland  Female
62                     Poland  Female
89                        NaN     NaN
215                       NaN     NaN
278  United States of America    Male
283                       NaN     NaN
297  United States of America    Male
306             United Kingdom    Male
340  United States of America    Male
348                       NaN     NaN
424  United States of America    Male
505             United Kingdom    Male
```

```
523                       NaN     NaN

                                 organization_name organization_city  \
18                                             NaN               NaN
62                             Sorbonne University             Paris
89                                             NaN               NaN
215                                            NaN               NaN
278  California Institute of Technology (Caltech)      Pasadena, CA
283                                            NaN               NaN
297                          University of Illinois        Urbana, IL
306                         University of Cambridge         Cambridge
340  California Institute of Technology (Caltech)      Pasadena, CA
348                                            NaN               NaN
424                          University of Illinois        Urbana, IL
505            MRC Laboratory of Molecular Biology         Cambridge
523                                            NaN               NaN

         organization_country  ISO  share_pct
18                        NaN  POL       0.25
62                     France  POL       1.00
89                        NaN  NaN       1.00
215                       NaN  NaN       1.00
278  United States of America  USA       1.00
283                       NaN  NaN       1.00
297  United States of America  USA       0.33
306            United Kingdom  GBR       1.00
340  United States of America  USA       1.00
348                       NaN  NaN       0.50
424  United States of America  USA       0.33
505            United Kingdom  GBR       0.25
523                       NaN  NaN       1.00
```

[29]:
```
col_subset = ['year', 'category', 'laureate_type', 'full_name']
multiple_winners[col_subset]
```

[29]:
```
     year   category laureate_type  \
18   1903    Physics    Individual
62   1911  Chemistry    Individual
89   1917      Peace  Organization
215  1944      Peace  Organization
278  1954  Chemistry    Individual
283  1954      Peace  Organization
297  1956    Physics    Individual
306  1958  Chemistry    Individual
340  1962      Peace    Individual
348  1963      Peace  Organization
424  1972    Physics    Individual
```

```
505  1980  Chemistry    Individual
523  1981      Peace  Organization

                                          full_name
18                        Marie Curie, née Sklodowska
62                        Marie Curie, née Sklodowska
89   Comité international de la Croix Rouge (Intern…
215  Comité international de la Croix Rouge (Intern…
278                                 Linus Carl Pauling
283  Office of the United Nations High Commissioner…
297                                       John Bardeen
306                                   Frederick Sanger
340                                 Linus Carl Pauling
348  Comité international de la Croix Rouge (Intern…
424                                       John Bardeen
505                                   Frederick Sanger
523  Office of the United Nations High Commissioner…
```

5 Number of Prizes per Category

[30]:
```
df_data.category.nunique()
```

[30]:
```
6
```

[31]:
```
df_data.category.value_counts()
```

[31]:
```
category
Medicine      222
Physics       216
Chemistry     186
Peace         135
Literature    117
Economics      86
Name: count, dtype: int64
```

[32]:
```
prizes_per_category = df_data.category.value_counts()
v_bar = px.bar(x = prizes_per_category.index,
               y= prizes_per_category.values,
               color = prizes_per_category.values,
               color_continuous_scale='Aggrnyl',
               title = 'Number of Prizes Awarded per Category')
v_bar.update_layout(xaxis_title = 'Nobel Prize Category',
                    coloraxis_showscale = False,
                    yaxis_title = 'Number of Prizes Awarded')
v_bar.show()
```

5.0.1 When was the first prize in the field of Economics awarded?

5.0.2 Who did the prize go to?

[33]:
```
df_data[df_data.category == 'Economics'].sort_values('year')[:3]
```

[33]:
```
     year   category                                                prize  \
393  1969  Economics  The Sveriges Riksbank Prize in Economic Scienc…
394  1969  Economics  The Sveriges Riksbank Prize in Economic Scienc…
402  1970  Economics  The Sveriges Riksbank Prize in Economic Scienc…

                                          motivation prize_share  \
393  "for having developed and applied dynamic mode…         1/2
394  "for having developed and applied dynamic mode…         1/2
402  "for the scientific work through which he has …         1/1

    laureate_type          full_name birth_date birth_city  \
393    Individual      Jan Tinbergen 1903-04-12  the Hague
394    Individual      Ragnar Frisch 1895-03-03       Oslo
402    Individual  Paul A. Samuelson 1915-05-15   Gary, IN

               birth_country     birth_country_current   sex  \
393              Netherlands               Netherlands  Male
394                   Norway                    Norway  Male
402  United States of America  United States of America  Male

                             organization_name organization_city  \
393         The Netherlands School of Economics         Rotterdam
394                           University of Oslo              Oslo
402  Massachusetts Institute of Technology (MIT)     Cambridge, MA

         organization_country  ISO  share_pct
393               Netherlands  NLD       0.50
394                    Norway  NOR       0.50
402  United States of America  USA       1.00
```

[34]:
```
cat_men_women = df_data.groupby(['category', 'sex'],as_index=False).
↪agg({'prize': pd.Series.count})
```

[35]:
```
cat_men_women
```

[35]:
```
      category     sex  prize
0    Chemistry  Female      7
1    Chemistry    Male    179
2    Economics  Female      2
3    Economics    Male     84
4   Literature  Female     16
5   Literature    Male    101
```

```
6     Medicine  Female     12
7     Medicine    Male    210
8        Peace  Female     17
9        Peace    Male     90
10     Physics  Female      4
11     Physics    Male    212
```

[36]:
```
cat_men_women.sort_values('prize', ascending=False, inplace=True)
```

[37]:
```
cat_men_women
```

[37]:
```
       category     sex  prize
11     Physics    Male    212
7     Medicine    Male    210
1    Chemistry    Male    179
5   Literature    Male    101
9        Peace    Male     90
3    Economics    Male     84
8        Peace  Female     17
4   Literature  Female     16
6     Medicine  Female     12
0    Chemistry  Female      7
10     Physics  Female      4
2    Economics  Female      2
```

[38]:
```
v_bar_split = px.bar(x = cat_men_women.category,
                     y = cat_men_women.prize,
                     color = cat_men_women.sex,
                     title = 'Number of Prizes Awarded per Category split by␣
 ↪Men and Women')
v_bar_split.update_layout(xaxis_title = 'Nobel Prize Category',
                          yaxis_title = 'Number of Prizes Awarded')
v_bar_split.show()
```

6 Number of Prizes Awarded Over Time

[39]:
```
df_data.year.value_counts()
```

[39]:
```
year
2001    15
2019    14
1996    13
2018    13
2000    13
        ..
1915     4
1924     3
```

```
1914     3
1918     2
1916     1
Name: count, Length: 117, dtype: int64
```

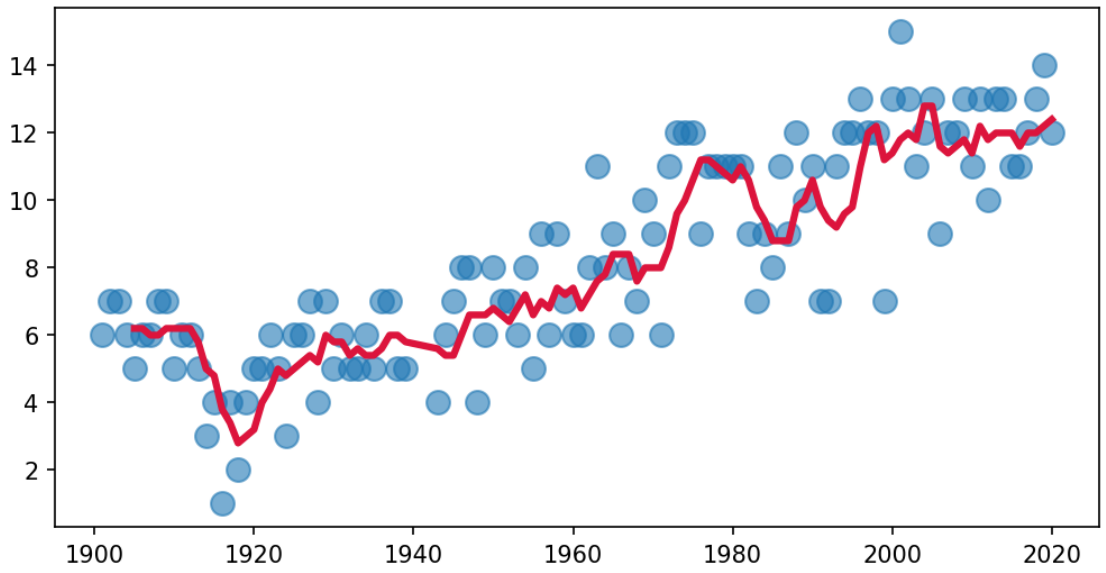
[40]:
```
prizes_per_year = df_data.groupby('year').count().prize
```

[41]:
```
prizes_per_year
```

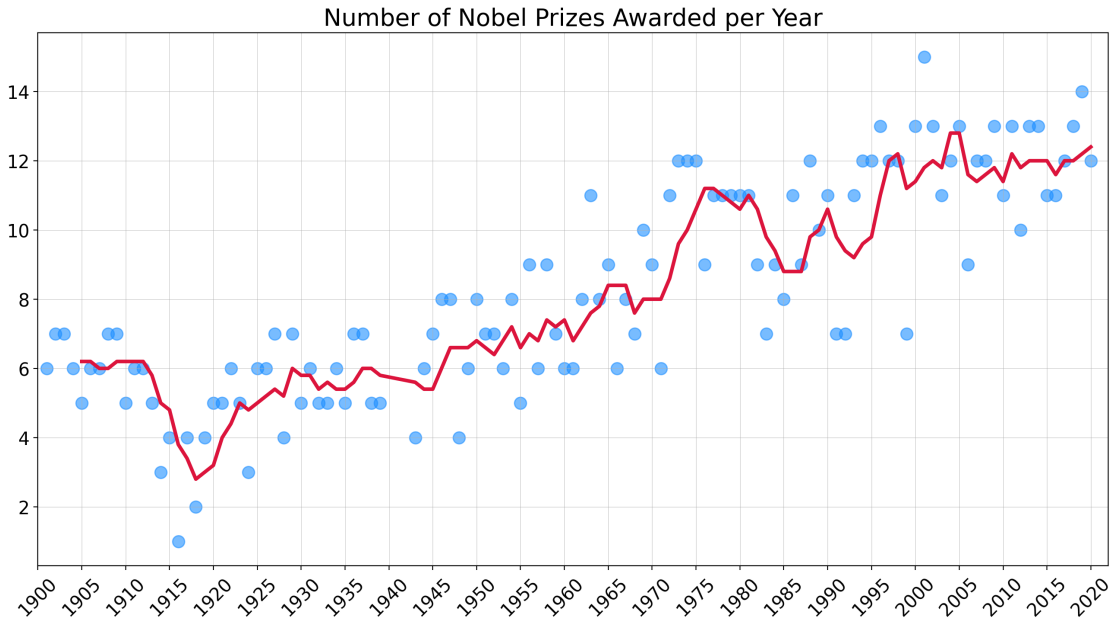
[41]:
```
year
1901     6
1902     7
1903     7
1904     6
1905     5
        ..
2016    11
2017    12
2018    13
2019    14
2020    12
Name: prize, Length: 117, dtype: int64
```

[42]:
```
moving_average = prizes_per_year.rolling(window = 5).mean()
```

[43]:
```
plt.figure(figsize = (8,4), dpi=150)

plt.scatter(x = prizes_per_year.index,
            y = prizes_per_year.values,
            alpha = 0.6,
            s = 100,)
plt.plot(prizes_per_year.index,
         moving_average.values,
         color = 'crimson',
         linewidth = 3,)
plt.show()
```



[44]:
```
plt.figure(figsize = (16,8), dpi=200)
plt.title('Number of Nobel Prizes Awarded per Year', fontsize = 20)


plt.yticks(fontsize = 15)
plt.xticks(ticks = np.arange(1900, 2021, step = 5),
           fontsize = 15,
           rotation = 45)

ax = plt.gca()

ax.set_xlim(1900, 2022)
ax.scatter(x = prizes_per_year.index,
            y = prizes_per_year.values,
            color = 'dodgerblue',
            alpha = 0.6,
            s = 100,)
ax.plot(prizes_per_year.index,
         moving_average.values,
         color = 'crimson',
         linewidth = 3,)
ax.grid(True, linewidth = 0.3)

plt.show()
```



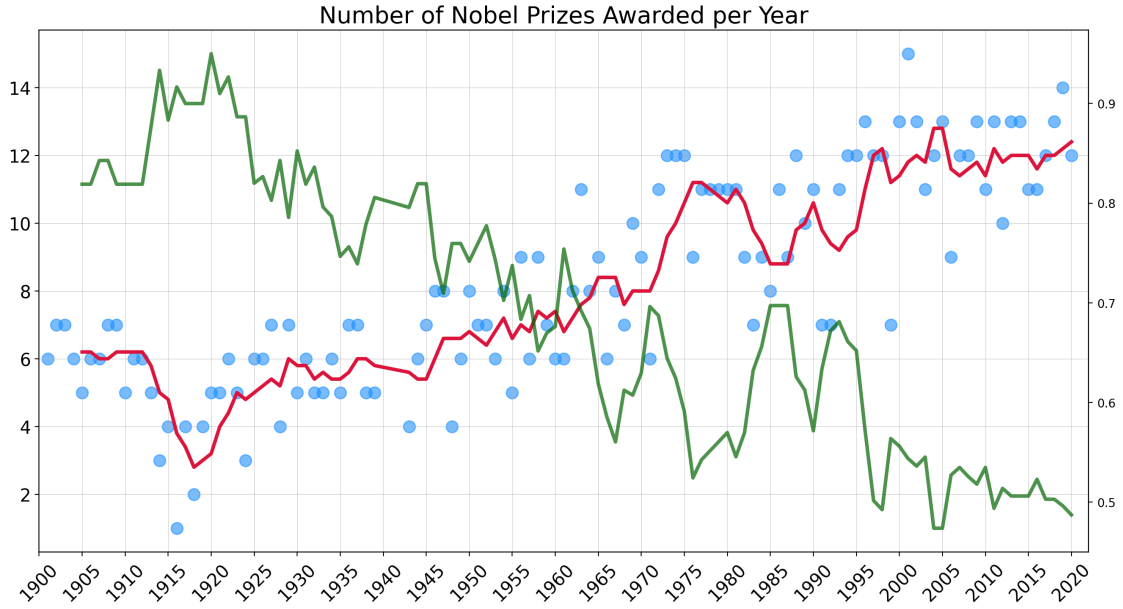
7 Are More Prizes Shared Than Before?

[45]:
```
yearly_average_share = df_data.groupby('year').agg({'share_pct':pd.Series.mean})
```

[46]:
```
share_moving_average = yearly_average_share.rolling(window = 5).mean()
```

[47]:
```
plt.figure(figsize = (16,8), dpi=200)
plt.title('Number of Nobel Prizes Awarded per Year', fontsize = 20)


plt.yticks(fontsize = 15)
plt.xticks(ticks = np.arange(1900, 2021, step = 5),
           fontsize = 15,
           rotation = 45)

ax1 = plt.gca()
ax2 = ax1.twinx()

ax1.set_xlim(1900, 2022)
ax1.scatter(x = prizes_per_year.index,
            y = prizes_per_year.values,
            color = 'dodgerblue',
            alpha = 0.6,
            s = 100,)
ax1.plot(prizes_per_year.index,
         moving_average.values,
```

```
          color = 'crimson',
          linewidth = 3,)

ax2.plot(prizes_per_year.index,
          share_moving_average.values,
          color = 'darkgreen',
          linewidth = 3,
          alpha = 0.7)

ax1.grid(True, linewidth = 0.3)

plt.show()
```



There is clearly an upward trend in the number of prizes being given out as more and more prizes are shared. Also, more prizes are being awarded from 1969 onwards because of the addition of the economics category. We also see that very few prizes were awarded during the first and second world wars. Note that instead of there being a zero entry for those years, we instead see the effect of the wars as missing blue dots.

8 The Countries with the Most Nobel Prizes

[48]:
```
top_countries = df_data.groupby(['birth_country_current'], as_index=False).
  ↪agg({'prize': pd.Series.count})
```

[49]:
```
top_countries.sort_values(by='prize', inplace=True)
top20_countries = top_countries[-20:]
```

[50]:
```
top20_countries
```

[50]:

	birth_country_current	prize
7	Belgium	9
31	Hungary	9
33	India	9
2	Australia	10
20	Denmark	12
54	Norway	12
13	China	12
51	Netherlands	18
3	Austria	18
39	Italy	19
68	Switzerland	19
11	Canada	20
61	Russia	26
40	Japan	27
57	Poland	27
67	Sweden	29
25	France	57
26	Germany	84
73	United Kingdom	105
74	United States of America	281

[51]:
```
h_bar = px.bar(x = top20_countries.prize,
               y = top20_countries.birth_country_current,
               orientation='h',
               color = top20_countries.prize,
               color_continuous_scale = 'Rainbow',
               title = 'Top 20 Countries by Number of Prizes')

h_bar.update_layout(xaxis_title = 'Number of Prizes',
                    yaxis_title = 'Country',
                    coloraxis_showscale = False)
h_bar.show()
```

9 Use a Choropleth Map to Show the Number of Prizes Won by Country

[52]:
```
df_countries = df_data.groupby(['birth_country_current', 'ISO'],␣
↪as_index=False).agg({'prize': pd.Series.count})
```

[53]:
```
df_countries
```

[53]:

	birth_country_current	ISO	prize
0	Algeria	DZA	2

```
1                   Argentina  ARG      4
2                   Australia  AUS     10
3                     Austria  AUT     18
4                  Azerbaijan  AZE      1
..                        …  …    …
74  United States of America  USA    281
75                  Venezuela  VEN      1
76                    Vietnam  VNM      1
77                      Yemen  YEM      1
78                   Zimbabwe  ZWE      1

[79 rows x 3 columns]
```

[55]:
```
world_map = px.choropleth(df_countries,
                          locations='ISO',
                          color='prize',
                          hover_name='birth_country_current',
                          color_continuous_scale=px.colors.sequential.matter)

world_map.update_layout(coloraxis_showscale=True,)

world_map.show()
```

[56]:
```
cat_country = df_data.groupby(['birth_country_current', 'category'],␣
 ↪as_index=False).agg({'prize': pd.Series.count})
cat_country.sort_values(by='prize', ascending=False, inplace=True)
```

[57]:
```
cat_country
```

[57]:
```
        birth_country_current   category  prize
204  United States of America   Medicine     78
206  United States of America    Physics     70
201  United States of America  Chemistry     55
202  United States of America  Economics     49
198            United Kingdom   Medicine     28
..                          …        …    …
97                       Iraq      Peace      1
99                    Ireland   Medicine      1
100                   Ireland    Physics      1
102                    Israel  Economics      1
210                  Zimbabwe      Peace      1

[211 rows x 3 columns]
```

[58]:
```
merged_df = pd.merge(cat_country, top20_countries, on='birth_country_current')
```

[59]:
```
merged_df
```

[59]:
```
         birth_country_current    category  prize_x  prize_y
0    United States of America    Medicine       78      281
1    United States of America     Physics       70      281
2    United States of America   Chemistry       55      281
3    United States of America   Economics       49      281
4    United States of America       Peace       19      281
..                        …         …      …      …
105                     India  Literature        2        9
106                     India    Medicine        2        9
107                     India   Chemistry        1        9
108                     India       Peace        1        9
109                     India     Physics        1        9

[110 rows x 4 columns]
```

[60]:
```
merged_df.columns = ['birth_country_current', 'category', 'cat_prize',␣
 ↪'total_prize']
merged_df.sort_values(by='total_prize', inplace=True)
```

[61]:
```
merged_df
```

[61]:
```
         birth_country_current   category  cat_prize  total_prize
109                     India    Physics          1            9
108                     India      Peace          1            9
88                    Belgium      Peace          3            9
89                    Belgium   Medicine          3            9
90                    Belgium  Chemistry          1            9
..                        …         …        …          …
4    United States of America      Peace         19          281
3    United States of America  Economics         49          281
2    United States of America  Chemistry         55          281
1    United States of America    Physics         70          281
0    United States of America   Medicine         78          281

[110 rows x 4 columns]
```

[62]:
```
cat_country_bar = px.bar(x=merged_df.cat_prize,
                         y=merged_df.birth_country_current,
                         color=merged_df.category,
                         orientation='h',
                         title='Top 20 Countries by Number of Prizes and␣
 ↪Category')

cat_country_bar.update_layout(xaxis_title='Number of Prizes',
                              yaxis_title='Country')
cat_country_bar.show()
```

Splitting the country bar chart by category allows us to get a very granular look at the data and

answer a whole bunch of questions. For example, we see is that the US has won an incredible proportion of the prizes in the field of Economics. In comparison, Japan and Germany have won very few or no economics prize at all. Also, the US has more prizes in physics or medicine alone than all of France's prizes combined. On the chart, we also see that Germany won more prizes in physics than the UK and that France has won more prizes in peace and literature than Germany, even though Germany has been awarded a higher total number of prizes than France.

10 In Which Categories are the Different Countries Winning Prizes?

[63]:
```
prizes_by_year = df_data.groupby(by=['birth_country_current', 'year'],␣
↪as_index=False).count()
prizes_by_year = prizes_by_year.sort_values('year')[['year',␣
↪'birth_country_current', 'prize']]
```

[64]:
```
prizes_by_year
```

[64]:

	year	birth_country_current	prize
118	1901	France	2
346	1901	Poland	1
159	1901	Germany	1
312	1901	Netherlands	1
440	1901	Switzerland	1
..	…	…	…
31	2019	Austria	1
221	2020	Germany	1
622	2020	United States of America	7
533	2020	United Kingdom	2
158	2020	France	1

[627 rows x 3 columns]

[65]:
```
cumulative_prizes = prizes_by_year.groupby(by=['birth_country_current','year']).
↪sum().groupby(level=[0]).cumsum()
cumulative_prizes.reset_index(inplace=True)
```

[66]:
```
cumulative_prizes
```

[66]:

	birth_country_current	year	prize
0	Algeria	1957	1
1	Algeria	1997	2
2	Argentina	1936	1
3	Argentina	1947	2
4	Argentina	1980	3
..	…	…	…
622	United States of America	2020	281
623	Venezuela	1980	1

```
624                  Vietnam  1973      1
625                    Yemen  2011      1
626                 Zimbabwe  1960      1

[627 rows x 3 columns]
```

[67]:
```
l_chart = px.line(cumulative_prizes,
                  x='year',
                  y='prize',
                  color='birth_country_current',
                  hover_name='birth_country_current')

l_chart.update_layout(xaxis_title='Year',
                      yaxis_title='Number of Prizes')

l_chart.show()
```

What we see is that the United States really started to take off after the Second World War which decimated Europe. Prior to that, the Nobel prize was pretty much a European affair. Very few laureates were chosen from other parts of the world. This has changed dramatically in the last 40 years or so. There are many more countries represented today than in the early days. Interestingly we also see that the UK and Germany traded places in the 70s and 90s on the total number of prizes won. Sweden being 5th place pretty consistently over many decades is quite interesting too. Perhaps this reflects a little bit of home bias?

10.0.1 Number of Prizes Won by Each Country Over Time

[68]:
```
top20_orgs = df_data.organization_name.value_counts()[:20]
top20_orgs.sort_values(ascending=True, inplace=True)
```

[69]:
```
top20_orgs
```

[69]:
```
organization_name
Sorbonne University                          7
Harvard Medical School                       7
Institut Pasteur                             7
London University                            7
Bell Laboratories                            8
Cornell University                           8
Yale University                              9
MRC Laboratory of Molecular Biology         10
University of Oxford                        12
Rockefeller University                      13
Max-Planck-Institut                         13
Princeton University                        15
California Institute of Technology (Caltech)  17
Columbia University                         17
University of Cambridge                     18
```

```
University of Chicago                          20
Massachusetts Institute of Technology (MIT)    21
Stanford University                            23
Harvard University                             29
University of California                       40
Name: count, dtype: int64
```

11 What are the Top Research Organisations?

[70]:
```
org_bar = px.bar(x = top20_orgs.values,
                 y = top20_orgs.index,
                 orientation='h',
                 color=top20_orgs.values,
                 color_continuous_scale=px.colors.sequential.haline,
                 title='Top 20 Research Institutions by Number of Prizes')

org_bar.update_layout(xaxis_title='Number of Prizes',
                      yaxis_title='Institution',
                      coloraxis_showscale=False)
org_bar.show()
```

12 Which Cities Make the Most Discoveries?

[71]:
```
top20_org_cities = df_data.organization_city.value_counts()[:20]
top20_org_cities.sort_values(ascending=True, inplace=True)
city_bar2 = px.bar(x = top20_org_cities.values,
                   y = top20_org_cities.index,
                   orientation='h',
                   color=top20_org_cities.values,
                   color_continuous_scale=px.colors.sequential.Plasma,
                   title='Which Cities Do the Most Research?')

city_bar2.update_layout(xaxis_title='Number of Prizes',
                        yaxis_title='City',
                        coloraxis_showscale=False)
city_bar2.show()
```

13 Where are Nobel Laureates Born?

[72]:
```
top20_cities = df_data.birth_city.value_counts()[:20]
top20_cities.sort_values(ascending=True, inplace=True)
city_bar = px.bar(x=top20_cities.values,
                  y=top20_cities.index,
                  orientation='h',
                  color=top20_cities.values,
```

```
                  color_continuous_scale=px.colors.sequential.Plasma,
                  title='Where were the Nobel Laureates Born?')

city_bar.update_layout(xaxis_title='Number of Prizes',
                       yaxis_title='City of Birth',
                       coloraxis_showscale=False)
city_bar.show()
```

14 Plotly Sunburst Chart Combining Country, City, and Organisation

[81]:
```
country_city_org = df_data.groupby(by=['organization_country',␣
↪'organization_city', 'organization_name'],
                                   as_index=False).agg({'prize': pd.Series.
↪count})

country_city_org = country_city_org.sort_values('prize', ascending=False)
```

[82]:
```
country_city_org
```

[82]:

	organization_country	organization_city	\
205	United States of America	Cambridge, MA	
280	United States of America	Stanford, CA	
206	United States of America	Cambridge, MA	
209	United States of America	Chicago, IL	
195	United States of America	Berkeley, CA	
..	…	…	
110	Japan	Sapporo	
111	Japan	Tokyo	
112	Japan	Tokyo	
113	Japan	Tokyo	
290	United States of America	Yorktown Heights, NY	

	organization_name	prize
205	Harvard University	29
280	Stanford University	23
206	Massachusetts Institute of Technology (MIT)	21
209	University of Chicago	20
195	University of California	19
..	…	…
110	Hokkaido University	1
111	Asahi Kasei Corporation	1
112	Kitasato University	1
113	Tokyo Institute of Technology	1
290	IBM Thomas J. Watson Research Center	1

```
[291 rows x 4 columns]
```

[83]:
```
burst = px.sunburst(country_city_org,
                    path=['organization_country', 'organization_city',␣
 ↪'organization_name'],
                    values='prize',
                    title='Where do Discoveries Take Place?',)

burst.update_layout(xaxis_title='Number of Prizes',
                    yaxis_title='City',
                    coloraxis_showscale=False)

burst.show()
```

15 Patterns in the Laureate Age at the Time of the Award

How Old Are the Laureates When the Win the Prize?

[84]:
```
birth_years = df_data.birth_date.dt.year
```

[85]:
```
birth_years
```

[85]:
```
0      1,852.00
1      1,839.00
2      1,854.00
3      1,822.00
4      1,828.00
         …
957    1,949.00
958         NaN
959    1,965.00
960    1,952.00
961    1,931.00
Name: birth_date, Length: 962, dtype: float64
```

[86]:
```
df_data['winning_age'] = df_data.year - birth_years
```

[95]:
```
df_data.winning_age
```

[95]:
```
0      49.00
1      62.00
2      47.00
3      79.00
4      73.00
         …
957    71.00
958      NaN
```

```
959   55.00
960   68.00
961   89.00
Name: winning_age, Length: 962, dtype: float64
```

15.0.1 Who were the oldest and youngest winners?

[97]:
```
display(df_data.nlargest(n=1, columns='winning_age'))
```

```
     year   category                                prize  \
937  2019  Chemistry  The Nobel Prize in Chemistry 2019

                                      motivation prize_share laureate_type  \
937  "for the development of lithium-ion batteries"         1/3    Individual

           full_name birth_date birth_city birth_country  \
937  John Goodenough 1922-07-25       Jena       Germany

    birth_country_current   sex    organization_name organization_city  \
937               Germany  Male  University of Texas         Austin TX

         organization_country  ISO  share_pct  winning_age
937  United States of America  DEU       0.33        97.00
```

[98]:
```
display(df_data.nsmallest(n=1, columns='winning_age'))
```

```
     year category                         prize  \
885  2014    Peace  The Nobel Peace Prize 2014

                                        motivation prize_share  \
885  "for their struggle against the suppression of…         1/2

    laureate_type         full_name birth_date birth_city birth_country  \
885    Individual  Malala Yousafzai 1997-07-12    Mingora      Pakistan

    birth_country_current     sex organization_name organization_city  \
885              Pakistan  Female               NaN                NaN

    organization_country  ISO  share_pct  winning_age
885                  NaN  PAK       0.50        17.00
```

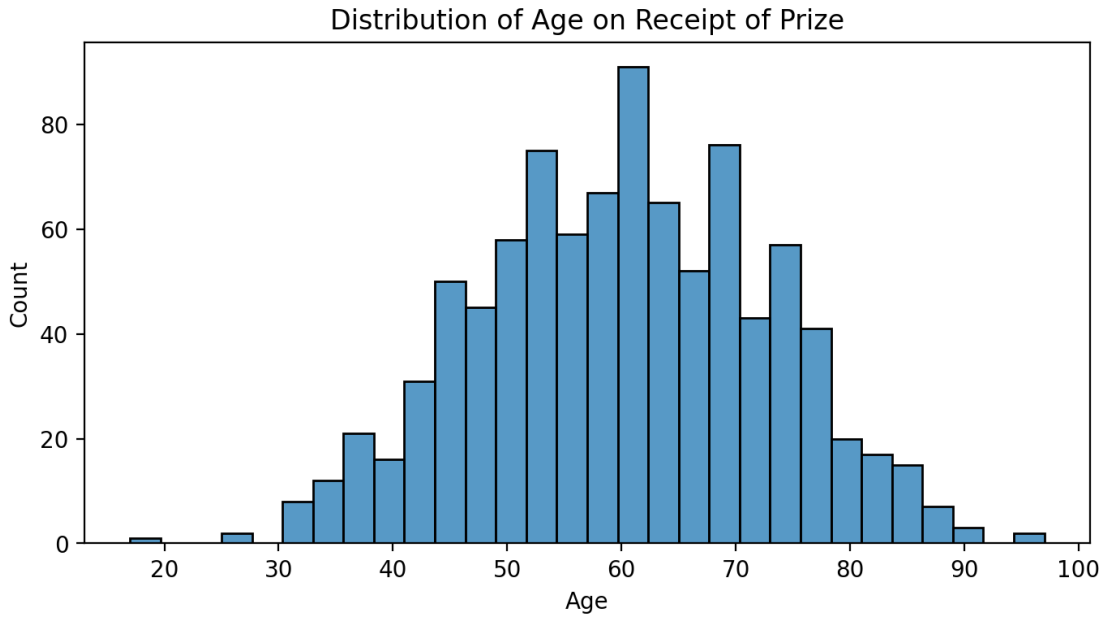
15.0.2 Descriptive Statistics for the Laureate Age at Time of Award

[99]:
```
df_data.winning_age.describe()
```

[99]:
```
count   934.00
mean     59.95
std      12.62
```

```
min      17.00
25%      51.00
50%      60.00
75%      69.00
max      97.00
Name: winning_age, dtype: float64
```

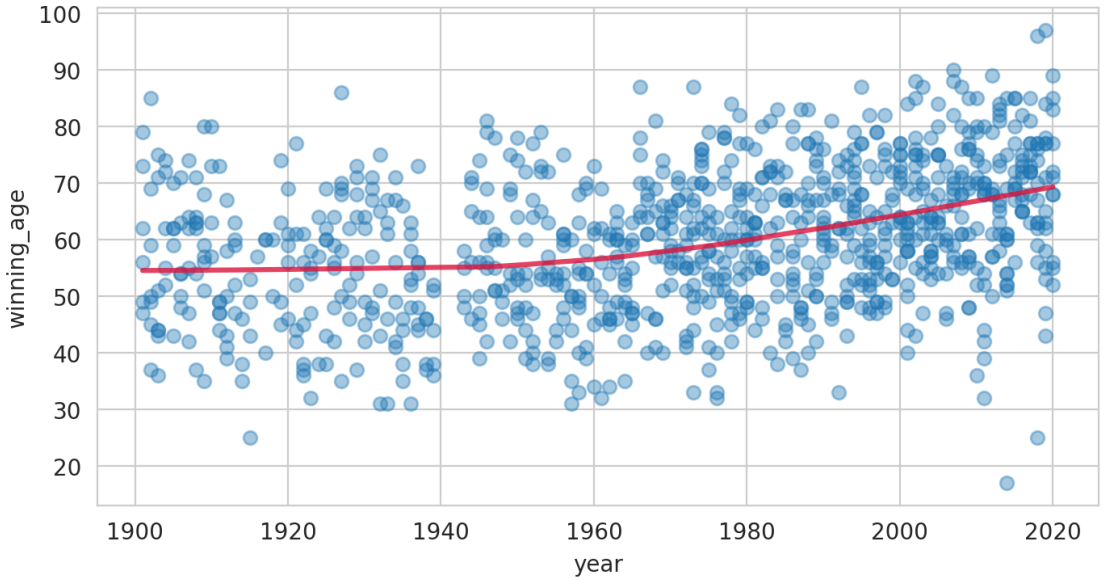
[105]:
```
plt.figure(figsize=(8, 4), dpi=200)

sns.histplot(data=df_data,
             x=df_data.winning_age,
             bins=30)

plt.xlabel('Age')
plt.title('Distribution of Age on Receipt of Prize')

plt.show()
```



15.0.3 Age at Time of Award throughout History

[113]:
```
plt.figure(figsize=(8,4), dpi=200)

with sns.axes_style("whitegrid"):
    sns.regplot(data=df_data,
                x='year',
                y='winning_age',
```

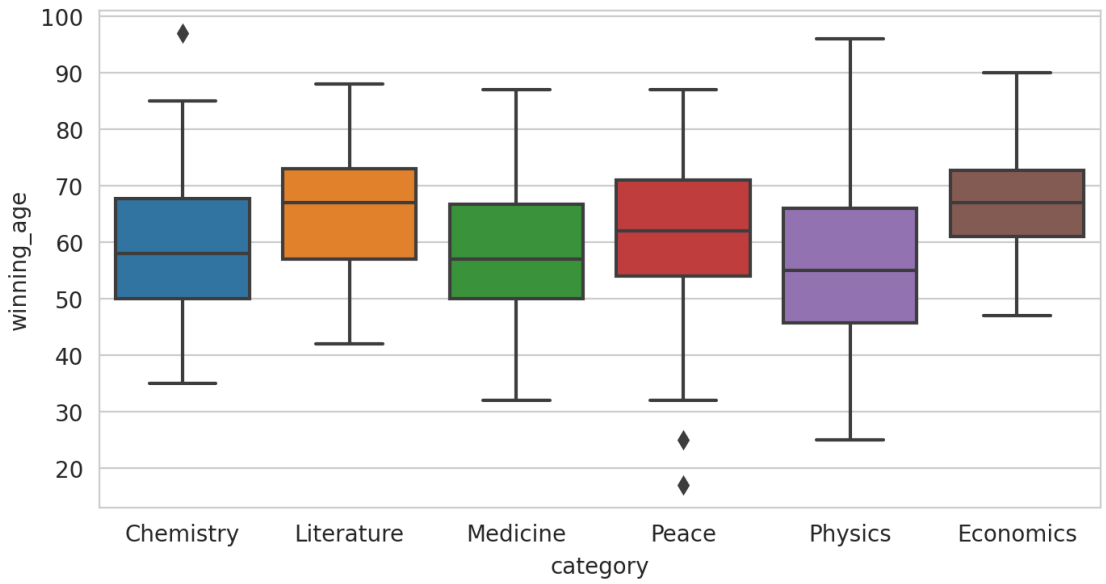
```
                lowess=True,
                scatter_kws = {'alpha': 0.4},
                line_kws={'color': 'crimson', 'alpha' : 0.8})
plt.show()
```



15.0.4 Winning Age Across the Nobel Prize Categories

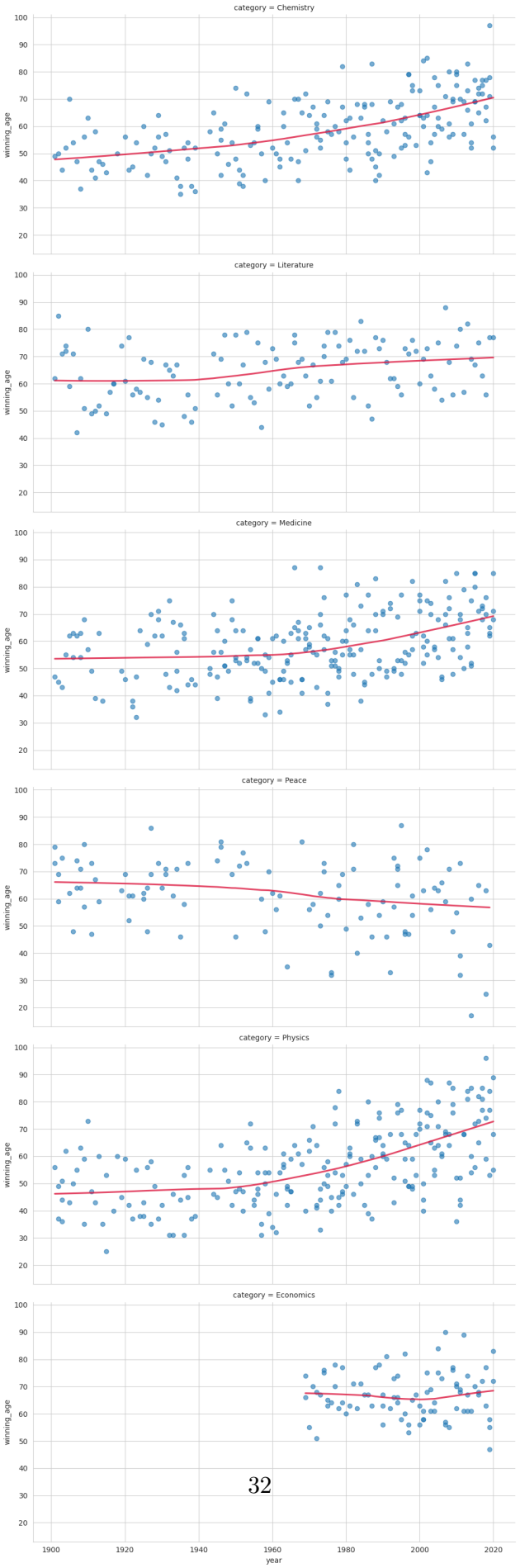
How does the age of laureates vary by category?

[115]:
```
plt.figure(figsize=(8,4), dpi=200)

with sns.axes_style("whitegrid"):
    sns.boxplot(data=df_data,
                x='category',
                y='winning_age')
plt.show()
```



Let's create 6 separate charts for each prize category using `.lmplot()`. * What are the winning age trends in each category? * Which category has the age trending up and which category has the age trending down? * Is this `.lmplot()` telling a different story from the `.boxplot()`? * Let's create another chart with Seaborn. This time let's use `.lmplot()` to put all 6 categories on the same chart using the `hue` parameter.

[117]:
```
with sns.axes_style('whitegrid'):
    sns.lmplot(data=df_data,
               x='year',
               y='winning_age',
               row = 'category',
               lowess=True,
               aspect=2,
               scatter_kws = {'alpha': 0.6},
               line_kws = {'color': 'crimson', 'alpha' : 0.8},)

plt.show()
```


category = Chemistry
category = Literature
category = Medicine
category = Peace
category = Physics
category = Economics
winning_age
year
100
90
80
70
60
50
40
30
20
1900
1920
1940
1960
1980
2000
2020

[123]:
```
with sns.axes_style("whitegrid"):
    sns.lmplot(data=df_data,
               x='year',
               y='winning_age',
               hue='category',
               lowess=True,
               aspect=2,
               scatter_kws={'alpha': 0.5},
               line_kws={'linewidth': 3, 'alpha' : 0.9})

plt.show()
```