History of Nobel Prize Winners

The Nobel Prize are five separate prizes that, according to Alfred Nobel's will of 1895, are awarded to "those who, during the preceding year, have conferred the greatest benefit to Mankind." Alfred Nobel was a Swedish chemist, engineer, and industrialist most famously known for the invention of dynamite. He died in 1896. In his will, he bequeathed all of his "remaining realisable assets" to be used to establish five prizes which became known as "Nobel Prizes." Nobel Prizes were first awarded in 1901.

Nobel Prizes are awarded in the fields of Physics, Chemistry, Physiology or Medicine, Literature, and Peace (Nobel characterized the Peace Prize as "to the person who has done the most or best to advance fellowship among nations, the abolition or reduction of standing armies, and the establishment and promotion of peace congresses"). In 1968, Sveriges Riksbank (Sweden's central bank) funded the establishment of the Prize in Economic Sciences in Memory of Alfred Nobel, to also be administered by the Nobel Foundation. Nobel Prizes are widely regarded as the most prestigious awards available in their respective fields.

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

In [2]:
```
nobel = pd.read_csv('nobel.csv')
```

Data Cleaning

In [3]:
```
nobel.head(10)
```

Out[3]:

	year	category	prize	motivation	prize_share	laureate_type	full_name	birth_date	birth_city	birth_country	birth_country_current
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
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
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
3	1901	Peace	The Nobel Peace Prize 1901	NaN	1/2	Individual	Frédéric Passy	1822-05-20	Paris	France	France
4	1901	Peace	The Nobel Peace Prize 1901	NaN	1/2	Individual	Jean Henry Dunant	1828-05-08	Geneva	Switzerland	Switzerland
5	1901	Physics	The Nobel Prize in Physics 1901	"in recognition of the extraordinary services ...	1/1	Individual	Wilhelm Conrad Röntgen	1845-03-27	Lennep (Remscheid)	Prussia (Germany)	Germany
6	1902	Chemistry	The Nobel Prize in Chemistry 1902	"in recognition of the extraordinary services ...	1/1	Individual	Hermann Emil Fischer	1852-10-09	Euskirchen	Prussia (Germany)	Germany
7	1902	Literature	The Nobel Prize in Literature 1902	"the greatest living master of the art of hist...	1/1	Individual	Christian Matthias Theodor Mommsen	1817-11-30	Garding	Schleswig (Germany)	Germany
8	1902	Medicine	The Nobel Prize in Physiology or Medicine 1902	"for his work on malaria, by which he has show...	1/1	Individual	Ronald Ross	1857-05-13	Almora	India	India
9	1902	Peace	The Nobel Peace Prize 1902	NaN	1/2	Individual	Charles Albert Gobat	1843-05-21	Tramelan	Switzerland	Switzerland

In [4]:
```
nobel.tail(10)
```

Out[4]:

	year	category	prize	motivation	prize_share	laureate_type	full_name	birth_date	birth_city	birth_country	birth_country_curr
952	2020	Economics	The Sveriges Riksbank Prize in Economic Scienc...	“for improvements to auction theory and invent...	1/2	Individual	Paul R. Milgrom	1948-04-20	Detroit, MI	United States of America	United State Ame
953	2020	Economics	The Sveriges Riksbank Prize in Economic Scienc...	“for improvements to auction theory and invent...	1/2	Individual	Robert B. Wilson	1937-05-16	Geneva, NE	United States of America	United State Ame
954	2020	Literature	The Nobel Prize in Literature 2020	“for her unmistakable poetic voice that with a...	1/1	Individual	Louise Glück	1943-04-22	New York, NY	United States of America	United State Ame
955	2020	Medicine	The Nobel Prize in Physiology or Medicine 2020	“for the discovery of Hepatitis C virus”	1/3	Individual	Charles M. Rice	1952-08-25	Sacramento, CA	United States of America	United State Ame
956	2020	Medicine	The Nobel Prize in Physiology or Medicine 2020	“for the discovery of Hepatitis C virus”	1/3	Individual	Harvey J. Alter	1935-09-12	New York, NY	United States of America	United State Ame
957	2020	Medicine	The Nobel Prize in Physiology or	“for the discovery of Hepatitis C	1/3	Individual	Michael Houghton	1949-07-02	NaN	United Kingdom	United Kingd

			Medicine 2020	virus”							
958	2020	Peace	The Nobel Peace Prize 2020	“for its efforts to combat hunger, for its con...	1/1	Organization	World Food Programme (WFP)	NaN	NaN	NaN	N
959	2020	Physics	The Nobel Prize in Physics 2020	“for the discovery of a supermassive compact o...	1/4	Individual	Andrea Ghez	1965-06-16	New York, NY	United States of America	United State Ame
960	2020	Physics	The Nobel Prize in Physics 2020	“for the discovery of a supermassive compact o...	1/4	Individual	Reinhard Genzel	1952-03-24	Bad Homburg vor der Höhe	Germany	Germ
961	2020	Physics	The Nobel Prize in Physics 2020	“for the discovery that black hole formation i...	1/2	Individual	Roger Penrose	1931-08-08	Colchester	United Kingdom	United Kingd

Check for Duplicates

In [5]:
```
print(f'Duplicates?\n{nobel.duplicated().values.any()}')
```

```
Duplicates?
False
```

Check for NaN values

In [6]:
```
print(f'NaN values among the data?\n{nobel.isna().values.any()}')
```

```
NaN values among the data?
True
```

In [7]:
```
nobel.isna().sum()
```

Out[7]:
```
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
```

NaN values for birth date are all organisations

In [8]:
```
col_subset = ['year','category', 'laureate_type',
              'birth_date','full_name', 'organization_name']
nobel.loc[nobel.birth_date.isna()][col_subset]
```

Out[8]:

	year	category	laureate_type	birth_date	full_name	organization_name
24	1904	Peace	Organization	NaN	Institut de droit international (Institute of ...	NaN
60	1910	Peace	Organization	NaN	Bureau international permanent de la Paix (Per...	NaN
89	1917	Peace	Organization	NaN	Comité international de la Croix Rouge (Intern...	NaN
200	1938	Peace	Organization	NaN	Office international Nansen pour les Réfugiés ...	NaN

215	1944	Peace	Organization	NaN	Comité international de la Croix Rouge (Intern...	NaN
237	1947	Peace	Organization	NaN	American Friends Service Committee (The Quakers)	NaN
238	1947	Peace	Organization	NaN	Friends Service Council (The Quakers)	NaN
283	1954	Peace	Organization	NaN	Office of the United Nations High Commissioner...	NaN
348	1963	Peace	Organization	NaN	Comité international de la Croix Rouge (Intern...	NaN
349	1963	Peace	Organization	NaN	Ligue des Sociétés de la Croix-Rouge (League o...	NaN
366	1965	Peace	Organization	NaN	United Nations Children's Fund (UNICEF)	NaN
399	1969	Peace	Organization	NaN	International Labour Organization (I.L.O.)	NaN
479	1977	Peace	Organization	NaN	Amnesty International	NaN
523	1981	Peace	Organization	NaN	Office of the United Nations High Commissioner...	NaN
558	1985	Peace	Organization	NaN	International Physicians for the Prevention of...	NaN
588	1988	Peace	Organization	NaN	United Nations Peacekeeping Forces	NaN
659	1995	Peace	Organization	NaN	Pugwash Conferences on Science and World Affairs	NaN
682	1997	Peace	Organization	NaN	International Campaign to Ban Landmines (ICBL)	NaN
703	1999	Peace	Organization	NaN	Médecins Sans Frontières	NaN
730	2001	Peace	Organization	NaN	United Nations (U.N.)	NaN
778	2005	Peace	Organization	NaN	International Atomic Energy Agency (IAEA)	NaN
788	2006	Peace	Organization	NaN	Grameen Bank	NaN
801	2007	Peace	Organization	NaN	Intergovernmental Panel on Climate Change (IPCC)	NaN
860	2012	Peace	Organization	NaN	European Union (EU)	NaN
873	2013	Peace	Organization	NaN	Organisation for the Prohibition of Chemical W...	NaN
897	2015	Peace	Organization	NaN	National Dialogue Quartet	NaN
919	2017	Peace	Organization	NaN	International Campaign to Abolish Nuclear Weap...	NaN
958	2020	Peace	Organization	NaN	World Food Programme (WFP)	NaN

NaN values for organisation_name

In [9]:
```
col_subset = ['year','category', 'laureate_type','full_name', 'organization_name']
nobel.loc[nobel.organization_name.isna()][col_subset]
```

Out[9]:

	year	category	laureate_type	full_name	organization_name
1	1901	Literature	Individual	Sully Prudhomme	NaN
3	1901	Peace	Individual	Frédéric Passy	NaN
4	1901	Peace	Individual	Jean Henry Dunant	NaN
7	1902	Literature	Individual	Christian Matthias Theodor Mommsen	NaN
9	1902	Peace	Individual	Charles Albert Gobat	NaN
...	...	...	...	...	...
932	2018	Peace	Individual	Nadia Murad	NaN
942	2019	Literature	Individual	Peter Handke	NaN
946	2019	Peace	Individual	Abiy Ahmed Ali	NaN
954	2020	Literature	Individual	Louise Glück	NaN
958	2020	Peace	Organization	World Food Programme (WFP)	NaN

255 rows × 5 columns

Convert Year and Birth Date to Datetime

In [10]:
```
nobel.birth_date = pd.to_datetime(nobel.birth_date)
```

Add a Column with the Prize Share as a Percentage

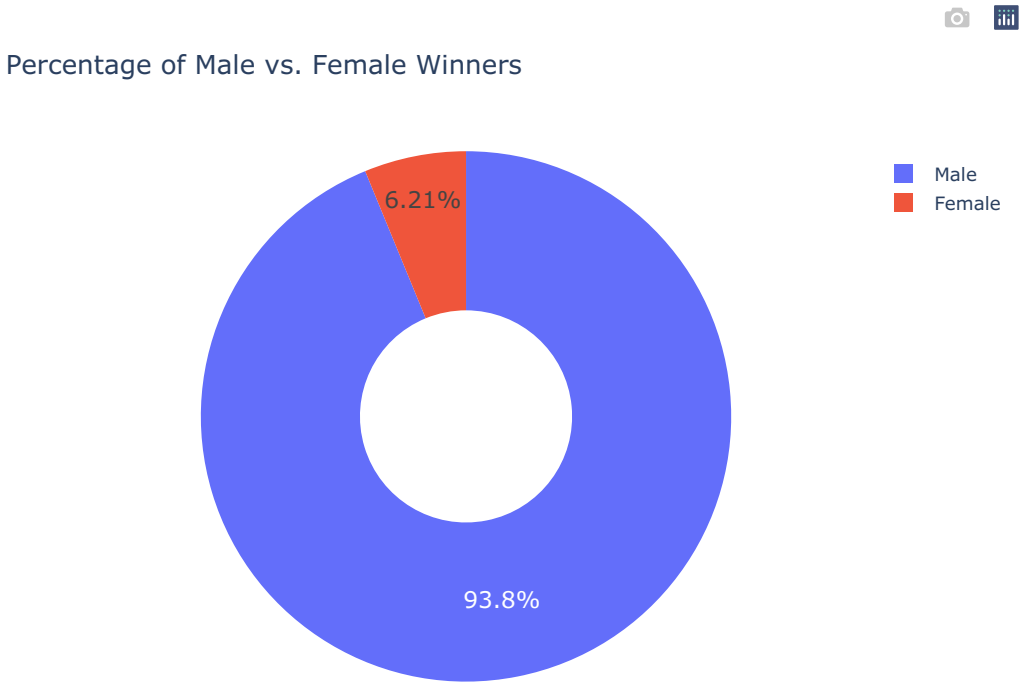
In [11]:
```
separated_values = nobel.prize_share.str.split('/', expand=True)
numerator = pd.to_numeric(separated_values[0])
denomenator = pd.to_numeric(separated_values[1])
nobel['share_pct'] = numerator / denomenator
```

In [12]:
```
nobel.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 962 entries, 0 to 961
Data columns (total 17 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
 16  share_pct              962 non-null    float64
dtypes: datetime64[ns](1), float64(1), int64(1), object(14)
memory usage: 127.9+ KB
```

Percentage of Male vs. Female Laureates

In [13]:
```
percentage = nobel.sex.value_counts()
fig = px.pie(labels=percentage.index,
             values=percentage.values,
             title="Percentage of Male vs. Female Winners",
             names=percentage.index,
             hole=0.4,)

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

fig.show()
```



Find repeat winners

In [14]:

In [14]:
```
single_winner = nobel.duplicated(subset=['full_name'], keep=False)
multiple_winners = nobel[single_winner]
print(f'There are {multiple_winners.full_name.nunique()}' \
      ' winners who weere awarded the prize more than once.')
```

```
There are 6 winners who weere awarded the prize more than once.
```

In [15]:
```
multiple_winners = nobel.groupby(by = 'full_name').filter(lambda x : x['year'].count() >= 2)
```

In [16]:
```
col_subset = ['year', 'category', 'laureate_type', 'full_name']
multiple_winners[col_subset]
```

Out[16]:

	year	category	laureate_type	full_name
18	1903	Physics	Individual	Marie Curie, née Sklodowska
62	1911	Chemistry	Individual	Marie Curie, née Sklodowska
89	1917	Peace	Organization	Comité international de la Croix Rouge (Intern...
215	1944	Peace	Organization	Comité international de la Croix Rouge (Intern...
278	1954	Chemistry	Individual	Linus Carl Pauling
283	1954	Peace	Organization	Office of the United Nations High Commissioner...
297	1956	Physics	Individual	John Bardeen
306	1958	Chemistry	Individual	Frederick Sanger
340	1962	Peace	Individual	Linus Carl Pauling
348	1963	Peace	Organization	Comité international de la Croix Rouge (Intern...
424	1972	Physics	Individual	John Bardeen
505	1980	Chemistry	Individual	Frederick Sanger
523	1981	Peace	Organization	Office of the United Nations High Commissioner...

Number of Prizes per Category

In [17]:
```
nobel.category.nunique()
```

Out[17]: 6

In [18]:
```
prizes_per_category = nobel.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='balance',
        title='Number of Prizes Awarded per Category')

v_bar.update_layout(xaxis_title='Nobel Prize Category',
                    coloraxis_showscale=False,
                    yaxis_title='Number of Prizes')
v_bar.show()
```

First prize in the feild of Economics awarded

In [19]:
```
nobel[nobel.category == 'Economics'].sort_values('year')[:3]
```

Out[19]:

	year	category	prize	motivation	prize_share	laureate_type	full_name	birth_date	birth_city	birth_country	birth_country_current	
393	1969	Economics	The Sveriges Riksbank Prize in Economic Scienc...	"for having developed and applied dynamic mode...	1/2	Individual	Jan Tinbergen	1903-04-12	the Hague	Netherlands	Netherlands	N
394	1969	Economics	The Sveriges Riksbank Prize in Economic Scienc...	"for having developed and applied dynamic mode...	1/2	Individual	Ragnar Frisch	1895-03-03	Oslo	Norway	Norway	N
402	1970	Economics	The Sveriges Riksbank Prize in Economic Scienc...	"for the scientific work through which he has ...	1/1	Individual	Paul A. Samuelson	1915-05-15	Gary, IN	United States of America	United States of America	N

Male and Female Winners by Category

In [20]:
```
cat_men_women = nobel.groupby(['category', 'sex'],
                              as_index=False).agg({'prize': pd.Series.count})
cat_men_women.sort_values('prize', ascending=False, inplace=True)
cat_men_women
```

Out[20]:

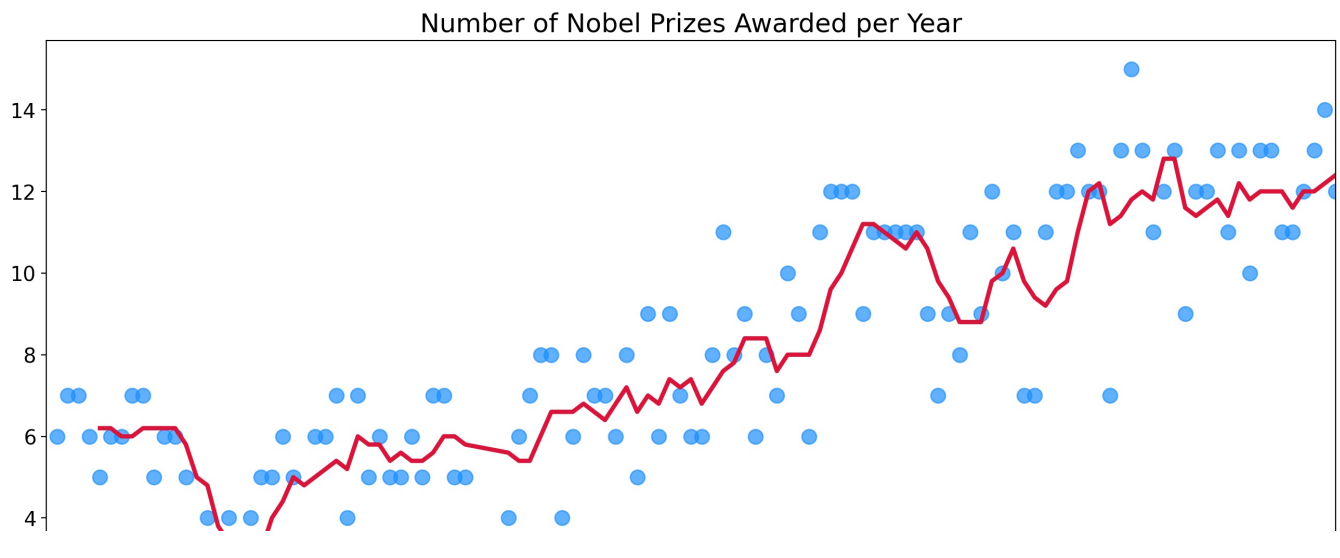
	category	sex	prize
11	Physics	Male	212
7	Medicine	Male	210
1	Chemistry	Male	179
5	Literature	Male	101
9	Peace	Male	90
3	Economics	Male	84
8	Peace	Female	17
4	Literature	Female	16
6	Medicine	Female	12
0	Chemistry	Female	7
10	Physics	Female	4
2	Economics	Female	2

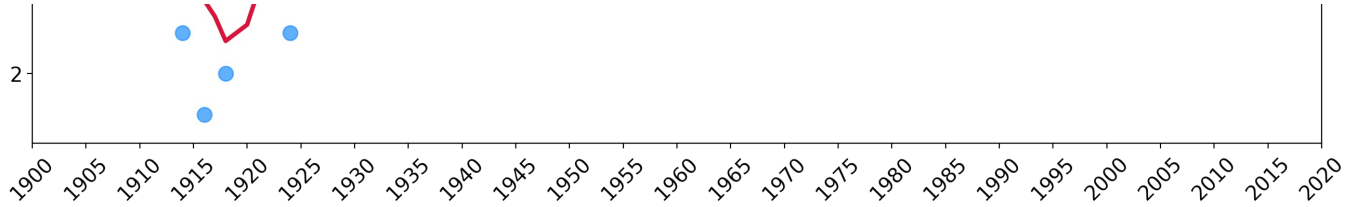
In [21]:
```
v_bar_split = px.bar(x = cat_men_women.category,
                     y = cat_men_women.prize,
                     color = cat_men_women.sex,
                     title='Number of Prizes Awarded per Category split by Men and Women')

v_bar_split.update_layout(xaxis_title='Nobel Prize Category',
                          yaxis_title='Number of Prizes')
v_bar_split.show()
```

Number of Prizes Awarded Over Time

In [22]:
```
prize_per_year = nobel.groupby(by='year').count().prize
moving_average = prize_per_year.rolling(window=5).mean()
```

In [23]:
```
plt.figure(figsize=(16,8), dpi=200)
plt.title('Number of Nobel Prizes Awarded per Year', fontsize=18)
plt.yticks(fontsize=14)
plt.xticks(ticks=np.arange(1900, 2021, step=5),
           fontsize=14,
           rotation=45)

ax = plt.gca()
ax.set_xlim(1900, 2020)

ax.scatter(x=prize_per_year.index,
           y=prize_per_year.values,
           c='dodgerblue',
           alpha=0.7,
           s=100,)

ax.plot(prize_per_year.index,
        moving_average.values,
        c='crimson',
        linewidth=3,)

plt.show()
```





The Countries with the Most Nobel Prizes

In [24]:
```
top_countries = nobel.groupby(['birth_country_current'],
                              as_index=False).agg({'prize': pd.Series.count})

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

Out[24]:

	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

In [25]:
```
h_bar = px.bar(x=top20_countries.prize,
               y=top20_countries.birth_country_current,
               orientation='h',
               color=top20_countries.prize,
               color_continuous_scale='Viridis',
               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()
```

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

In [26]:
```
df_countries = nobel.groupby(['birth_country_current', 'ISO'],
                             as_index=False).agg({'prize': pd.Series.count})
df_countries.sort_values('prize', ascending=False)
```

Out[26]:

	birth_country_current	ISO	prize
74	United States of America	USA	281
73	United Kingdom	GBR	105
26	Germany	DEU	84
25	France	FRA	57
67	Sweden	SWE	29
...	...	...	...
32	Iceland	ISL	1
47	Madagascar	MDG	1
34	Indonesia	IDN	1
36	Iraq	IRQ	1
78	Zimbabwe	ZWE	1

79 rows × 3 columns

In [27]:
```
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()
```

Cities with the Most Discoveries

In [28]:
```
top20_org_cities = nobel.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()
```

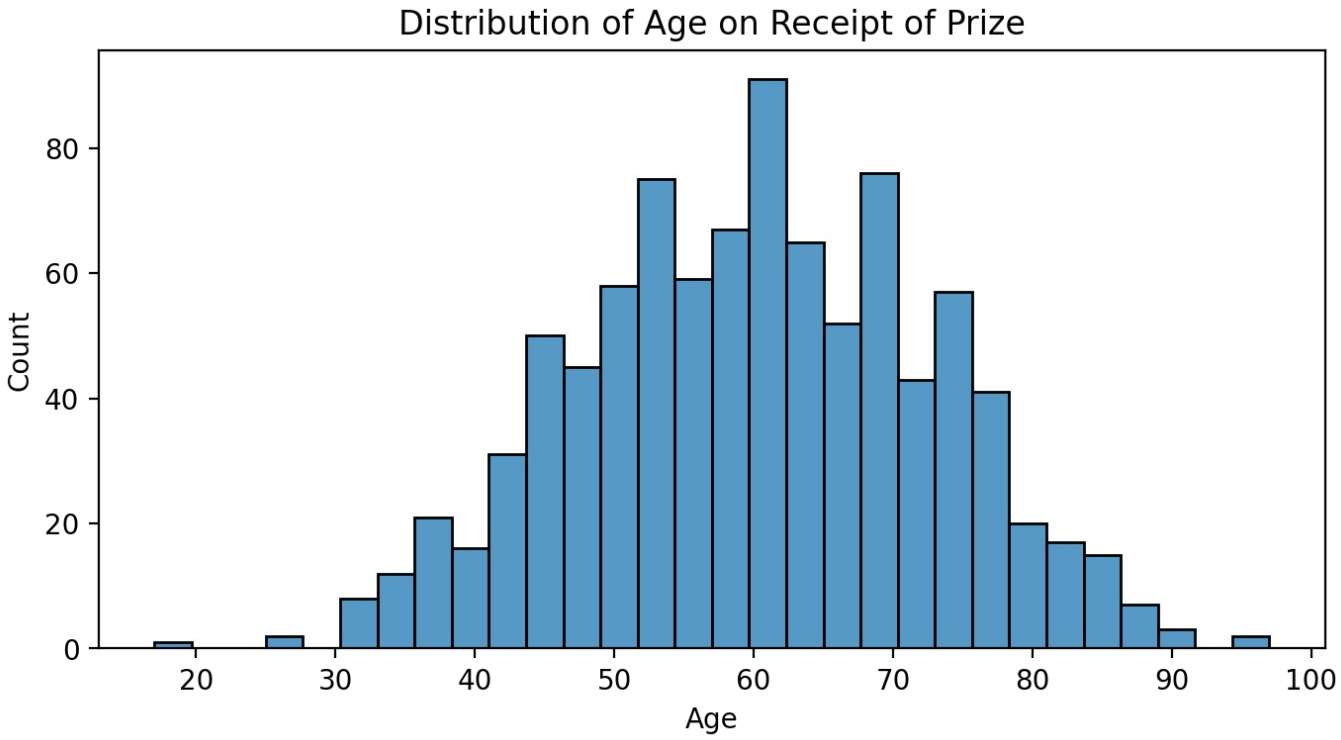
Descriptive Statistics for the Laureate Age at Time of Award

In [29]:
```
birth_years = nobel.birth_date.dt.year
birth_years
nobel['winning_age'] = nobel.year - birth_years
nobel.winning_age
nobel.winning_age.describe()
```

Out[29]:
```
count    934.000000
mean      59.948608
std       12.617785
min       17.000000
25%       51.000000
50%       60.000000
75%       69.000000
max       97.000000
Name: winning_age, dtype: float64
```

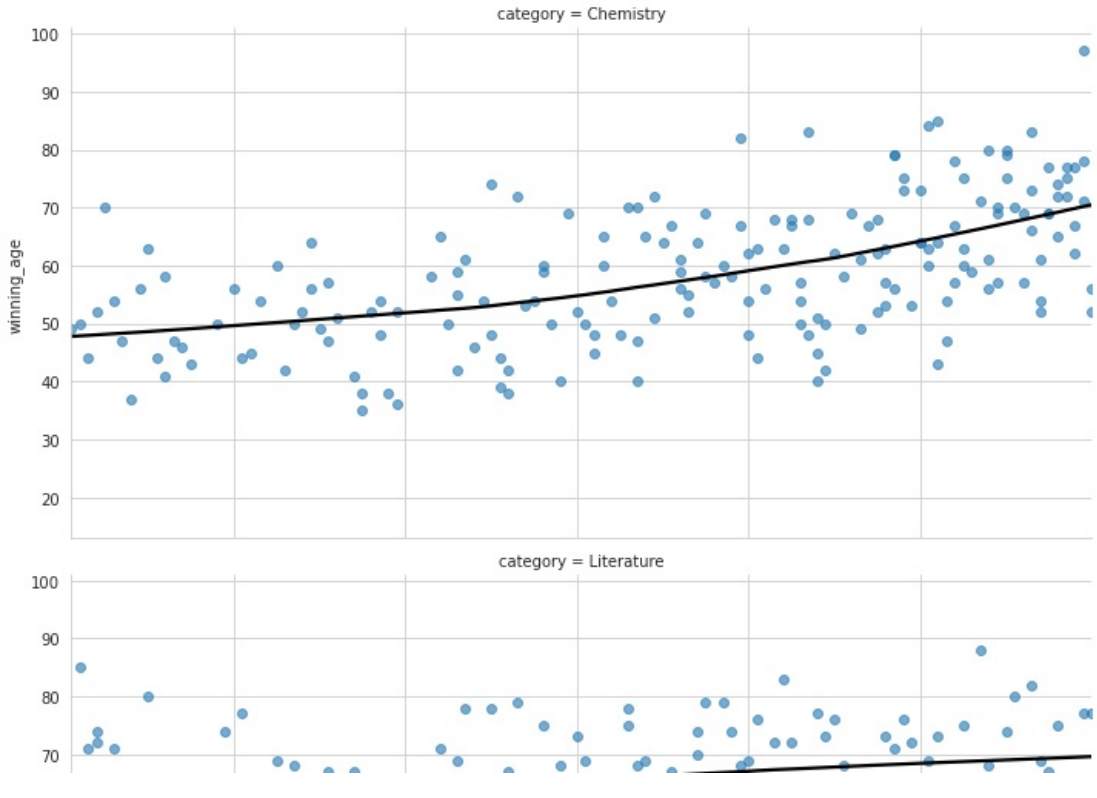
In [30]:

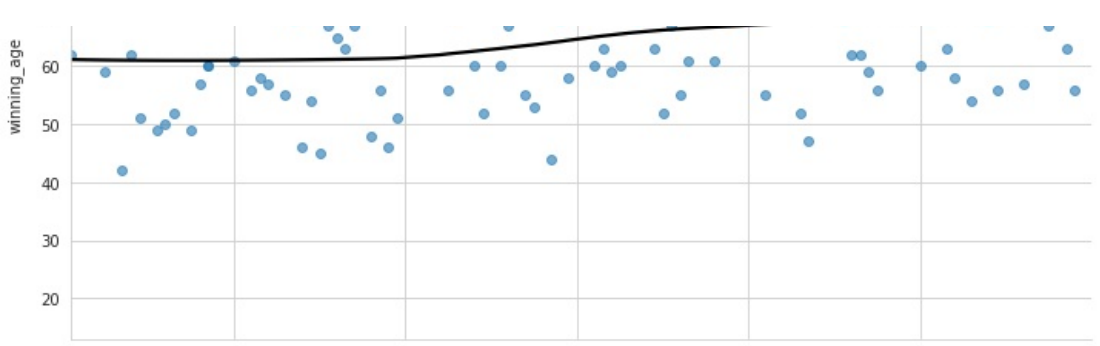
```
plt.figure(figsize=(8, 4), dpi=200)
sns.histplot(data=nobel,
             x=nobel.winning_age,
             bins=30)
plt.xlabel('Age')
plt.title('Distribution of Age on Receipt of Prize')
plt.show()
```

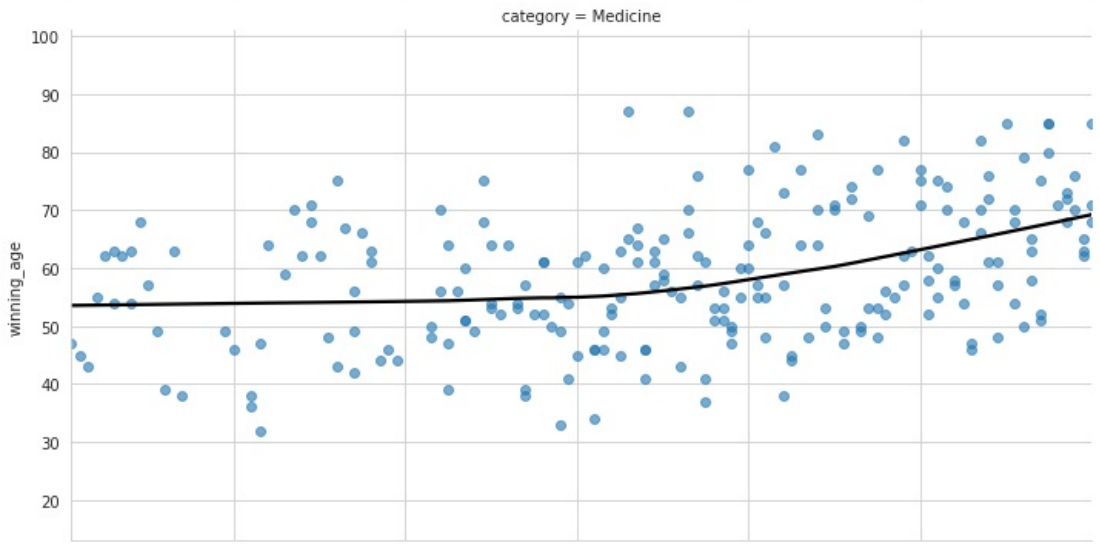


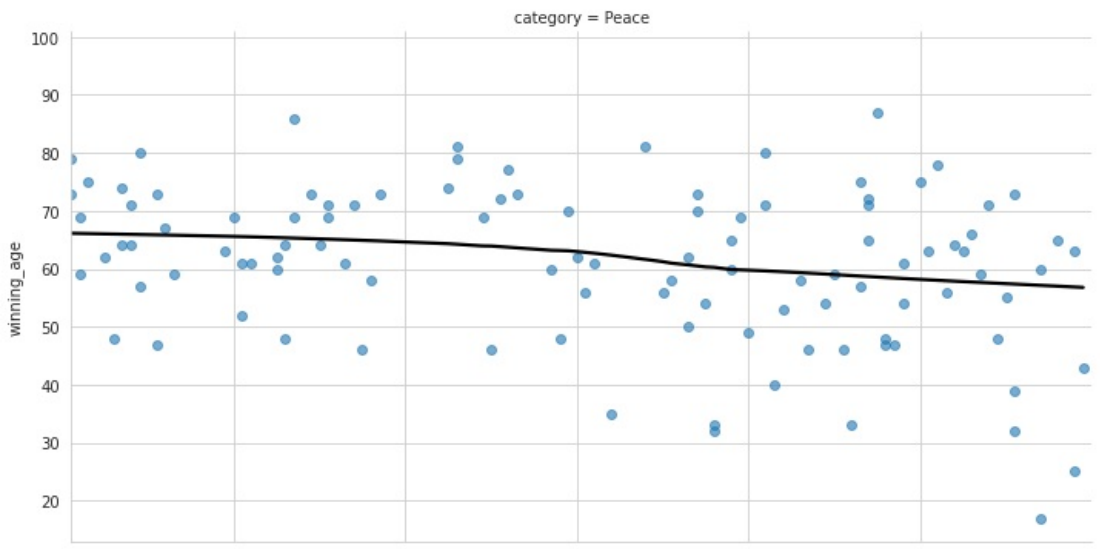
In [31]:

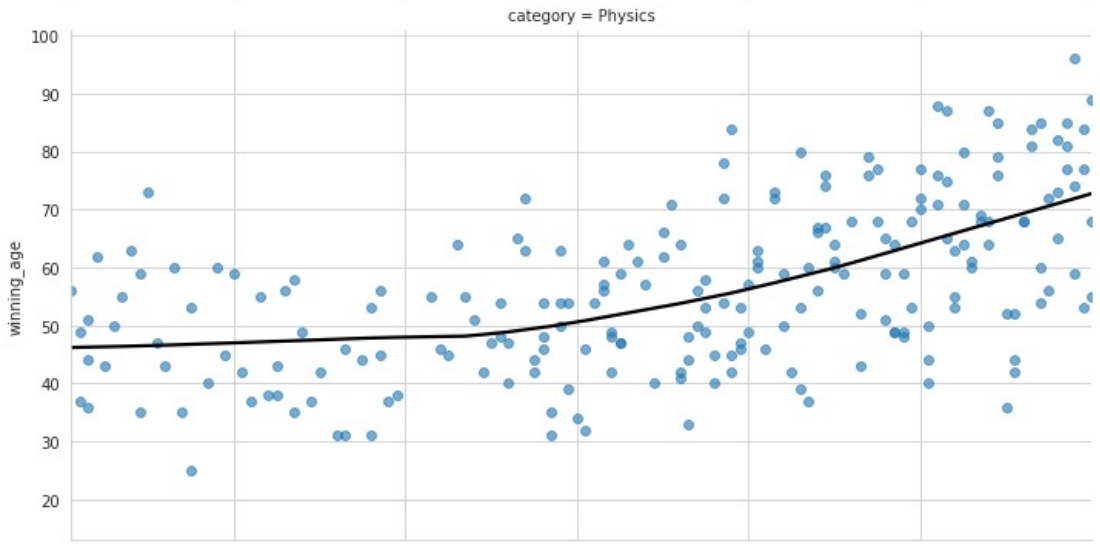
```
with sns.axes_style('whitegrid'):
    sns.lmplot(data=nobel,
               x='year',
               y='winning_age',
               row = 'category',
               lowess=True,
               aspect=2,
               scatter_kws = {'alpha': 0.6},
               line_kws = {'color': 'black'},)

plt.show()
```

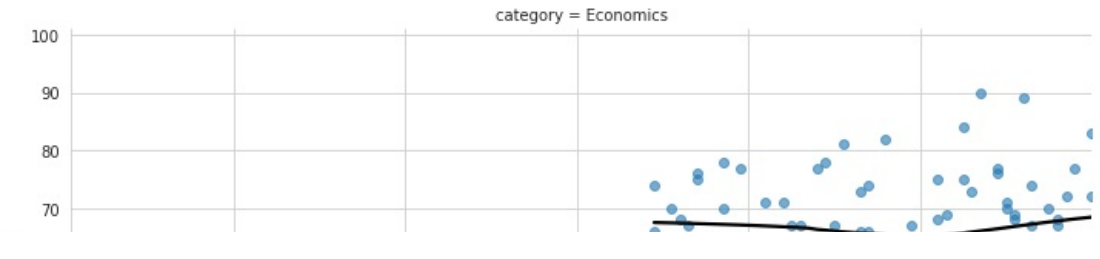



winning_age
60
50
40
30
20


category = Medicine
winning_age
100
90
80
70
60
50
40
30
20


category = Peace
winning_age
100
90
80
70
60
50
40
30
20


category = Physics
winning_age
100
90
80
70
60
50
40
30
20


category = Economics
100
90
80
70

In [32]:
```
display(nobel.nlargest(n=1, columns='winning_age'))
display(nobel.nsmallest(n=1, columns='winning_age'))
```

	year	category	prize	motivation	prize_share	laureate_type	full_name	birth_date	birth_city	birth_country	birth_country_current
937	2019	Chemistry	The Nobel Prize in Chemistry 2019	“for the development of lithium-ion batteries”	1/3	Individual	John Goodenough	1922-07-25	Jena	Germany	Germany

	year	category	prize	motivation	prize_share	laureate_type	full_name	birth_date	birth_city	birth_country	birth_country_current	se
885	2014	Peace	The Nobel Peace Prize 2014	"for their struggle against the suppression of...	1/2	Individual	Malala Yousafzai	1997-07-12	Mingora	Pakistan	Pakistan	Femal

The first 3 Women to Win the Nobel Prize

In [33]:
```
nobel [nobel.sex == 'Female'].sort_values('year', ascending=True)[:3]
```

Out[33]:

	year	category	prize	motivation	prize_share	laureate_type	full_name	birth_date	birth_city	birth_country	birth_country_current	
18	1903	Physics	The Nobel Prize in Physics 1903	"in recognition of the extraordinary services ...	1/4	Individual	Marie Curie, née Sklodowska	1867-11-07	Warsaw	Russian Empire (Poland)	Poland	F
29	1905	Peace	The Nobel Peace Prize 1905	NaN	1/1	Individual	Baroness Bertha Sophie Felicita von Suttner, n...	1843-06-09	Prague	Austrian Empire (Czech Republic)	Czech Republic	F
51	1909	Literature	The Nobel Prize in Literature 1909	"in appreciation of the lofty idealism, vivid ...	1/1	Individual	Selma Ottilia Lovisa Lagerlöf	1858-11-20	Mårbacka	Sweden	Sweden	F