The Nobel Prize has been among the most prestigious international awards since 1901. Each year, awards are bestowed in chemistry, literature, physics, physiology or medicine, economics, and peace. In addition to the honor, prestige, and substantial prize money, the recipient also gets a gold medal with an image of Alfred Nobel (1833 - 1896), who established the prize.

The Nobel Foundation has made a dataset available of all prize winners from the outset of the awards from 1901 to 2023. The dataset used in this project is from the Nobel Prize API and is available in the `nobel.csv` file in the `data` folder.

In this project, you'll get a chance to explore and answer several questions related to this prizewinning data. And we encourage you then to explore further questions that you're interested in!# Python Code

```
# Loading in required libraries
library(tidyverse)
library(readr)
library(ggplot2)
```

```
# 1. Load the dataset and find the most common gender and birth country
nobel = read_csv("data/nobel.csv")

head(nobel)

gender_nobel <- nobel %>%
    count(sex, sort = TRUE)

top_gender <- gender_nobel$sex[1]

country_nobel <- nobel %>%
    count(birth_country, sort = TRUE)

top_country <- country_nobel$birth_country[1]
```

```
Warning message:
“One or more parsing issues, call `problems()` on your data frame for details,
e.g.:
  dat <- vroom(...)
  problems(dat)”
Rows: 1000 Columns: 18
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr  (14): category, prize, motivation, prize_share, laureate_type, full_nam...
dbl   (2): year, laureate_id
date  (2): birth_date, death_date

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

	...	c...	...	prize	...	motivation	...	pri...	...	lau...	...	laurea
1	1901	Chemistry		The Nobel Prize in Chemistry 1901		"in recognition of the extraordinary services ...		1/1		160		Individ
2	1901	Literature		The Nobel Prize in Literature 1901		"in special recognition of his poetic composit...		1/1		569		Individ
3	1901	Medicine		The Nobel Prize in Physiology or Medicine 19...		"for his work on serum therapy, especially its...		1/1		293		Individ
4	1901	Peace		The Nobel Peace Prize 1901		null		1/2		462		Individ
5	1901	Peace		The Nobel Peace Prize 1901		null		1/2		463		Individ
6	1901	Physics		The Nobel Prize in Physics 1901		"in recognition of the extraordinary services ...		1/1		1		Individ

Rows: 6

```
# 2. Identify the decade with the highest proportion of US-born winners
decade_nobel_data <- nobel %>%
    mutate(decade = floor(year/10)*10)

# decade_nobel_data

decade_nobel <- decade_nobel_data %>%
    count(decade, sort = TRUE)

decade_usa_nobel <- decade_nobel_data %>%
    filter(birth_country == "United States of America") %>%
    count(decade, sort = TRUE)

# decade_usa_nobel

decade_usa_nobel_prop <- decade_usa_nobel %>%
    left_join(decade_nobel, "decade", suffix = c("_usa", "_all")) %>%
    mutate(usa_prop = n_usa/n_all) %>%
    arrange(desc(usa_prop))
```

```r
# decade_usa_nobel_prop

max_decade_usa <- decade_usa_nobel_prop$decade[1]
max_decade_usa
```

```
2000
```

```r
# 3. Find the decade and category with the highest proportion of female laureates
decade_woman_nobel <- decade_nobel_data %>%
    filter(sex == "Female") %>%
    count(decade, sort = TRUE)

decade_woman_nobel_prop <- decade_woman_nobel %>%
    left_join(decade_nobel, "decade", suffix = c("_woman", "_all")) %>%
    mutate(woman_prop = n_woman/n_all) %>%
    arrange(desc(woman_prop))

# decade_woman_nobel_prop

category_nobel <- nobel %>%
    count(category, sort = TRUE)

category_woman_nobel <- nobel %>%
    filter(sex == "Female") %>%
    count(category, sort = TRUE)

category_woman_nobel_prop <- category_woman_nobel %>%
    left_join(category_nobel, "category", suffix = c("_woman", "_all")) %>%
    mutate(woman_prop = n_woman/n_all) %>%
    arrange(desc(woman_prop))

# category_woman_nobel_prop

max_female_list <- list(decade_woman_nobel_prop$decade, category_woman_nobel_prop$category)
```

```r
# 4. Find first woman to win a Nobel Prize
first_woman_nobel <- nobel %>%
    filter(sex == "Female")

# nobel
# first_woman_nobel

first_woman_name <- first_woman_nobel$full_name[1]
#first_woman_name <- word(first_woman_name, 1)
# first_woman_name

first_woman_category <- first_woman_nobel$category[1]
```

```r
# 5. Determine repeat winners
repeats <- nobel %>%
    count(full_name, sort = TRUE) %>%
    filter(n > 1)
#repeats
```