CS612 – Data Exploration Visualization – Team Nobel

Dhanalakshmi Rajagopalan

Derek Bugg

Yesuf Abdissa

23rd August 2020

TP02

Table of Contents

[1. Introduction 2](#_Toc49087650)

[1.1 Overview 2](#_Toc49087651)

[1.2 Purpose 3](#_Toc49087652)

[1.3 Scope 3](#_Toc49087653)

[2. Prerequisites 3](#_Toc49087654)

[3. Planning 3](#_Toc49087655)

[3.1 Tasks 3](#_Toc49087656)

[4. Finding Insights from Data 4](#_Toc49087657)

[5. Structuring Nobel Prize Dataset 5](#_Toc49087658)

[5.1 Background and Context 5](#_Toc49087659)

[5.2 Integration with dataset and visualization 5](#_Toc49087660)

[6. Conclusion [TP03] 13](#_Toc49087661)

[6.1 The Future of Data visualization 13](#_Toc49087662)

[6.2 Lesson Learned 13](#_Toc49087663)

[Reference 14](#_Toc49087664)

# Introduction

## Overview

Data Visualization is a discipline to understand data by providing a graphical representation of data. It shows a great understanding of what the data represents and how the data can be used. The essential goal of data visualization is to communicate information clearly and effectively via graphical means.

The Nobel prize is the most prestigious well-known awards which are given to scholars and scientist in literature, physics, chemistry, medicine, peace, and economics. In 1901, the first Nobel prize was handed out at Nybroviken (Sweden). In this project, we will explore pattern and trends in over 100 years’ worth of Nobel prize winners.

## Purpose

The purpose of this project to consume different skill sets such as data visualization, data manipulation, and data cleaning for collecting data to discover interesting information about Nobel Prize winners.

## Scope

The scope of this project is to interpret the Nobel Prize dataset into a visual way of communication.

# Prerequisites

Before starting the project implementation, we should understand the concepts:

* Data Manipulation with Python (Pandas)
* Data visualization with Matplotlib or Seaborn
* Git repository
* [Dataset](https://www.kaggle.com/nobelfoundation/nobel-laureates)

# Planning

In data visualization, the planning phase is essential to make sure the outcome of the project is a compelling visualization. Our team wants to narrate the following exciting facts as tasks:

## Tasks

* who gets the Nobel Prize?
* the typical age of getting the prize,
* Has anybody gotten it twice?
* A typical Nobel Prize winner’s gender
* repeat laureates,
* oldest and youngest winners, and
* dominance by individual countries
* age differences between prize categories
* why there are some controversies

# Finding Insights from Data

Visualizations allow us to discover patterns and insights which can be used to tell stories. The insights can represent various things like visualizing, analyzing and interpreting what we see, documenting insight steps and transforming datasets.

Visualizing allow us to show the dataset in various ways such as charts, tables, maps, and graphs. For example, if we want to get “who gets the Nobel Prize”, we have to collect the data by going through history from 1901 to 2016. Instead, we can get the information from the dataset to convey the number of possible Nobel prizes handed out between 1901 to 2016.

Analyzing and interpreting data helps us to ask questions ourselves like Is it what we expected to visualize, Are there any exciting pattern? What does this represent in the context of data? These questions help us to find meaning within the relevant information from the dataset.

Documenting and insight help us to understand our expectation and assumptions about our dataset.

Transforming dataset allows us to explore more patterns and findings on the Nobel Prize dataset, which helps us to perform further inspection and analysis of data. The action can be done by aggregating data points into a single group, filtering and removing unrelated data.

# Structuring Nobel Prize Dataset

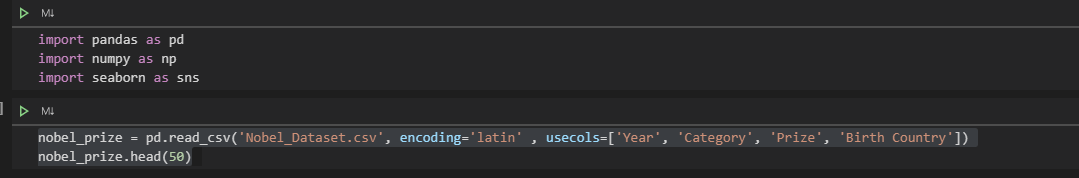
## Background and Context

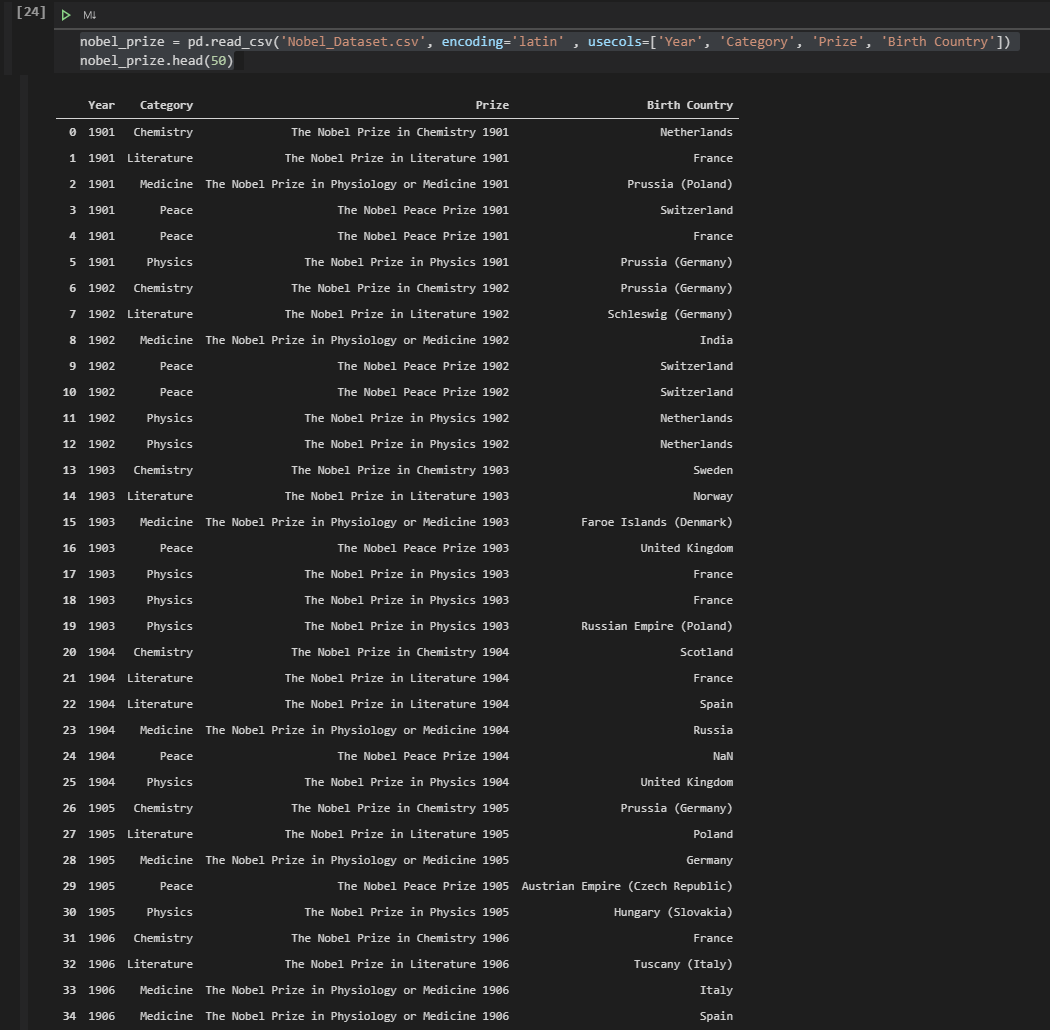
The dataset shows all the award’s inceptions in 1901 through October 2016. However, In 2019, Nobelprize.org opened [API](https://app.swaggerhub.com/apis/NobelMedia/NobelMasterData/2#/default/get_nobelPrizes) (Application Program Interface) for developers to know the latest information about the Nobel Prizes and the Nobel Laureates. Initially, we are going to visualize data with the current [dataset](https://www.kaggle.com/nobelfoundation/nobel-laureates?select=archive.csv). Later, It will be updated with the current information.

## Integration with dataset and visualization

* dominance by the USA

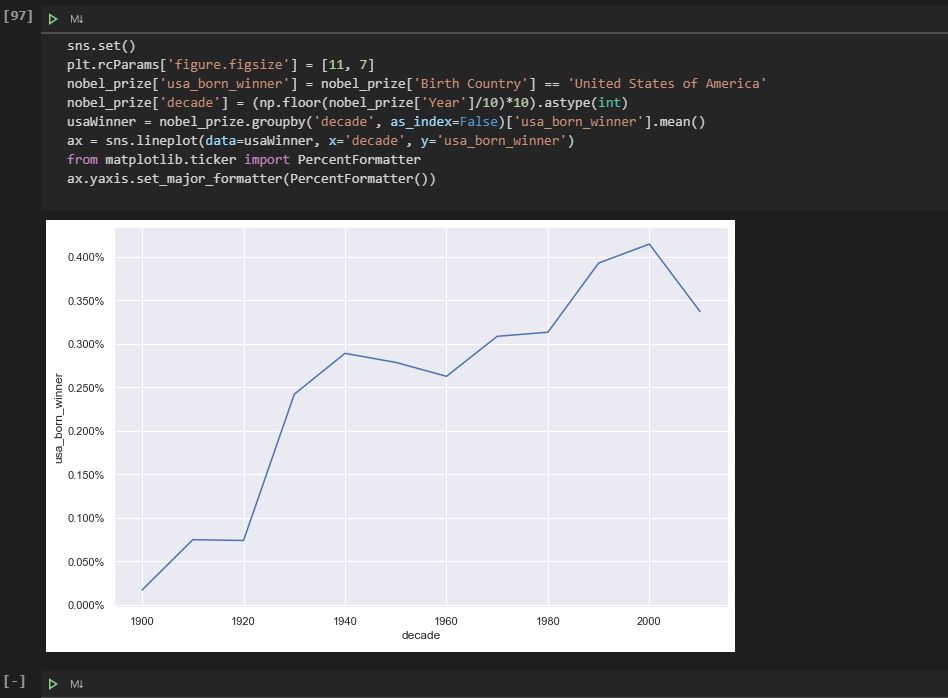
According to Wiki, the most top three countries are the United States, the United Kingdoms and Germany. When we look at the data prize chart from 1901 to 1906, all the winners were from Europe. Well, we are going to find out when did the United States of America start to dominate the Nobel prize charts.





The diagram below shows that the USA start dominates from 1920.





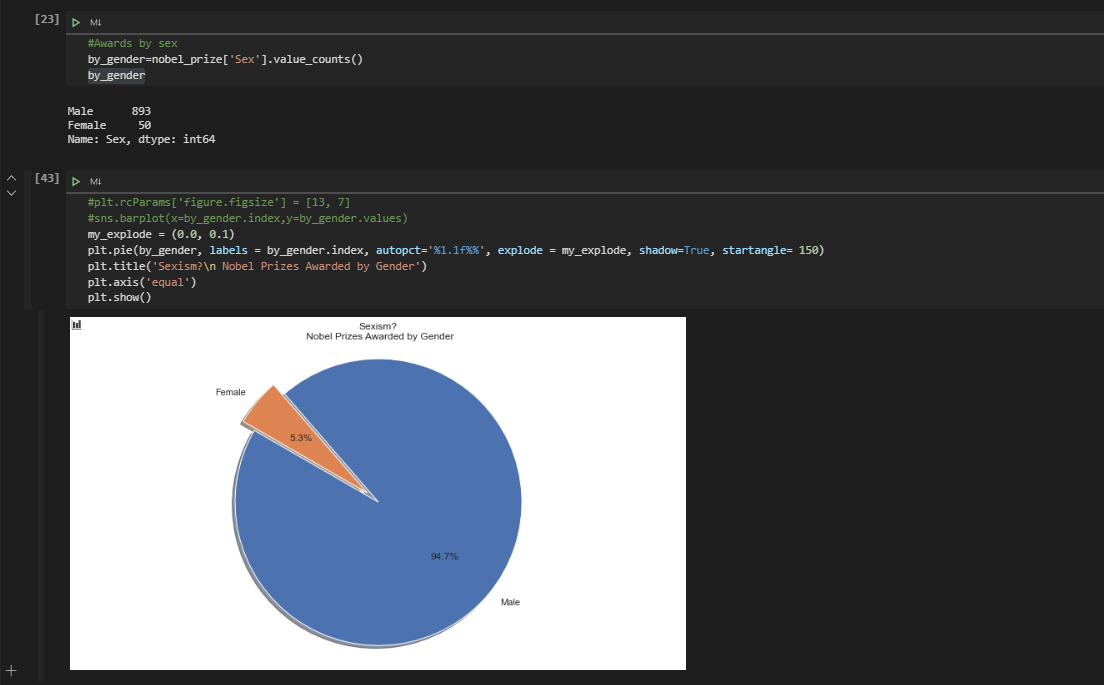
* age differences between prize categories [TP03]

It will be completed on our next submission.

* why there are some controversies[TP03]

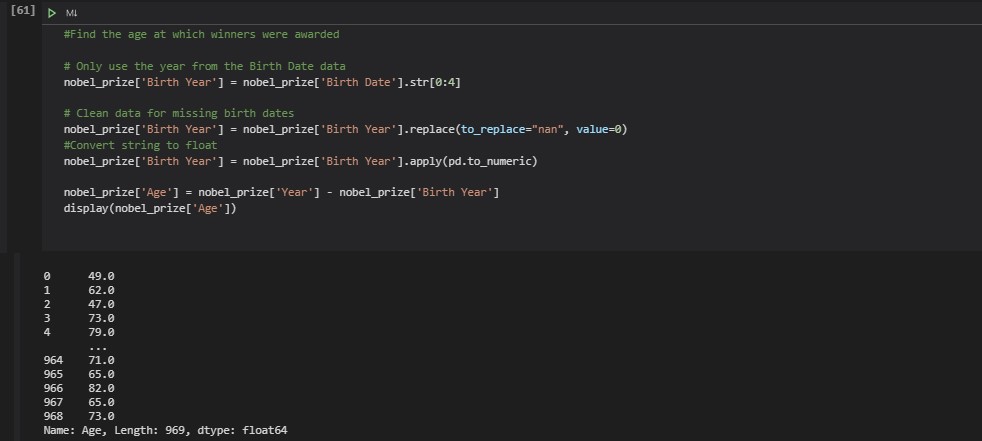
It will be completed on our next submission.

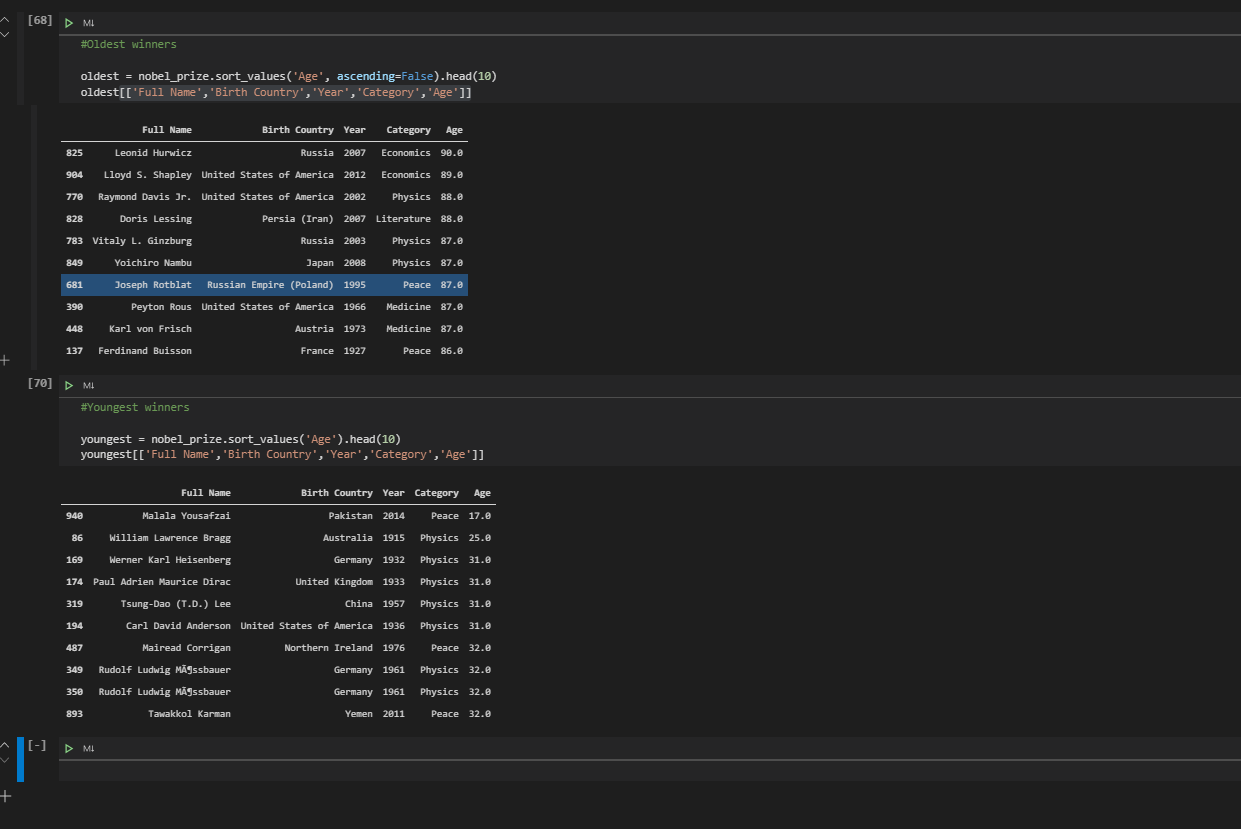
* a typical Nobel Prize winner’s gender



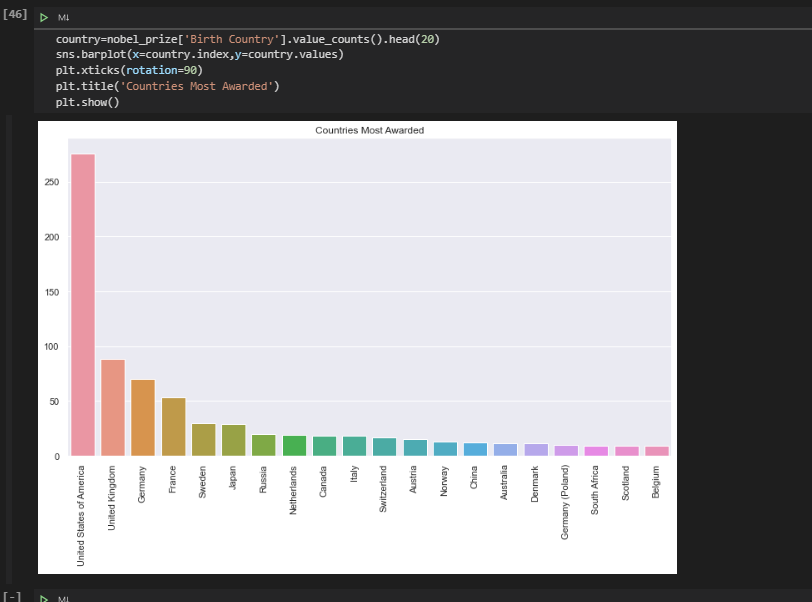
Summary and discussion of gender data to be completed for TP03.

* oldest and youngest winners





Summary and analysis Next Submission [TP03]

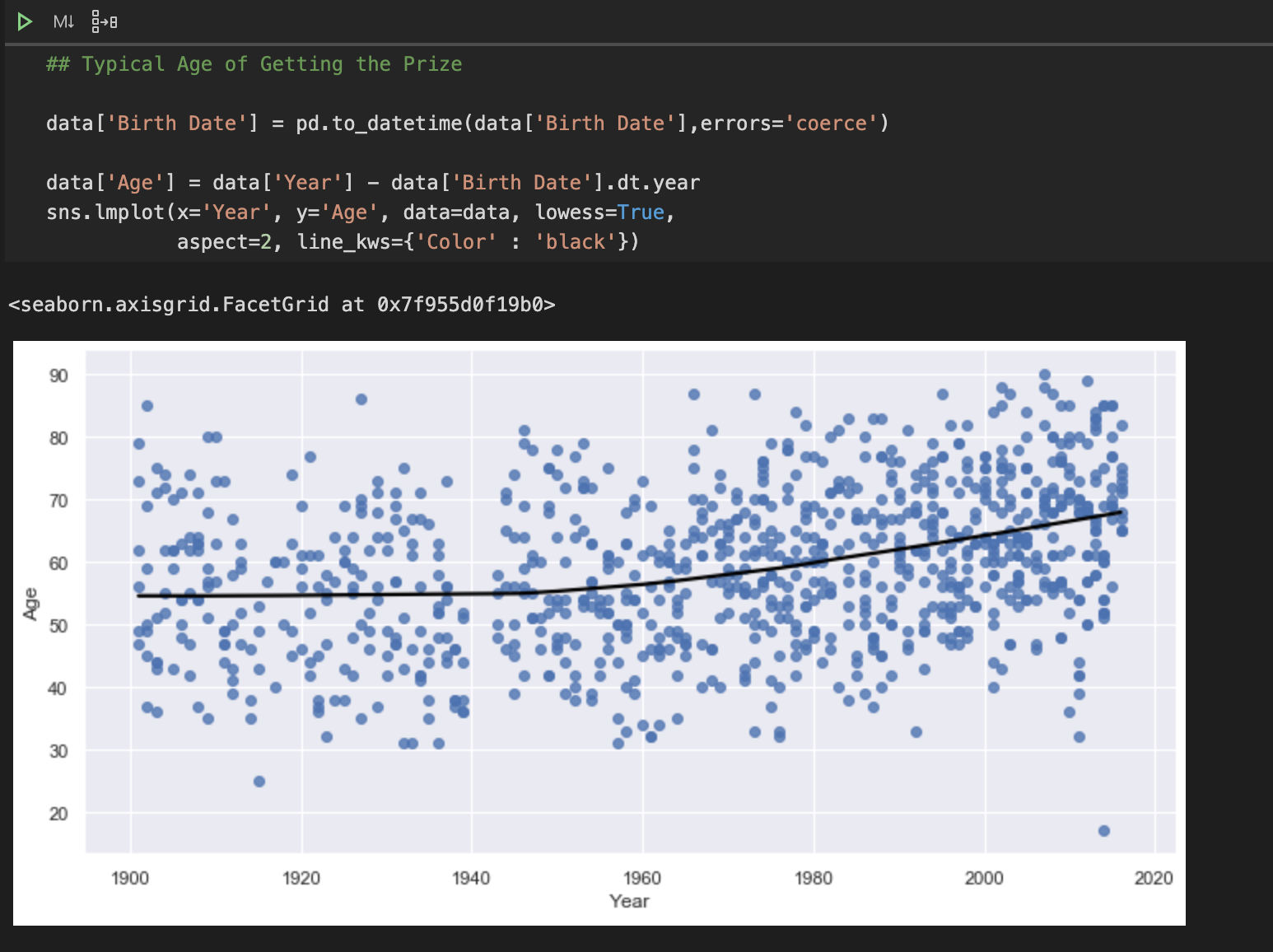


* who gets the Nobel Prize? [TP03]

Summary and analysis Next Submission

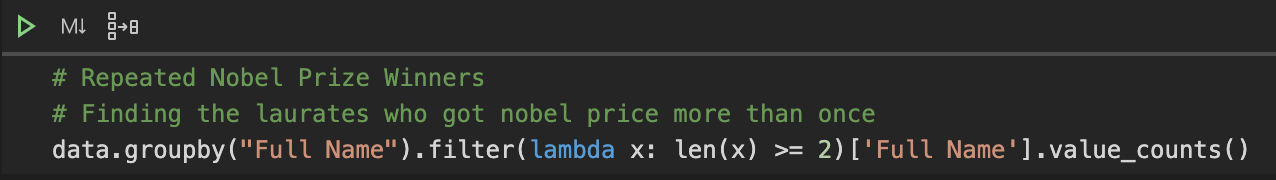
* the typical age of getting the prize,[TP03]

Summary and analysis Next Submission



* Has anybody gotten it twice?

qSummary and analysis Next Submission [TP03]





# Conclusion [TP03]

## The Future of Data visualization

## Lesson Learned

# Reference

Developer zone. (n.d.). Retrieved August 20, 2020, from <https://www.nobelprize.org/about/developer-zone-2/>

The Nobel Prize, (n. d.). Nobel Prize Facts. Retrieve from <https://www.nobelprize.org/prizes/facts/nobel-prize-facts/>

Pandas, (n. D)