Analytics report for Presidents Longevity

Sylvia Xu

December 1st 2022

Summary

This report analyzed the life span of U.S. presidents. Information is sourced from PresidentUSA.net with a total of 45 presidents. According to the data, The average life span of U.S. presidents(passed away) is 71 years, within a range from 94 to 46. The life span of U.S. presidents generally follows a normal distribution, but with more longer life individuals.

Assumptions regarding dataset:

1. Due to the availibility and responsibility of the resource, I assume the dataset is free of error (such as misspellings).
2. To simplify the outcomes and improve readbility, I purposely ignore any timezone related differences since it will have little impact towards the analysis.

```python
In [1]: #Import some analytic related libraries
        import pandas as pd
        import numpy as np
        from datetime import datetime
        import matplotlib.pyplot as plt
```

In [2]:
```python
#Read data as df
df = pd.read_csv("U.S. Presidents Birth and Death Information - Sheet1.csv")

#Get to know the data type and shape
print(df.dtypes)
print(df.shape)
df
```

```
PRESIDENT          object
BIRTH DATE         object
BIRTH PLACE        object
DEATH DATE         object
LOCATION OF DEATH  object
dtype: object
(46, 5)
```

Out[2]:

	PRESIDENT	BIRTH DATE	BIRTH PLACE	DEATH DATE	LOCATION OF DEATH
0	George Washington	Feb 22, 1732	Westmoreland Co., Va.	Dec 14, 1799	Mount Vernon, Va.
1	John Adams	Oct 30, 1735	Quincy, Mass.	July 4, 1826	Quincy, Mass.
2	Thomas Jefferson	Apr 13, 1743	Albemarle Co., Va.	July 4, 1826	Albemarle Co., Va.
3	James Madison	Mar 16, 1751	Port Conway, Va.	June 28, 1836	Orange Co., Va.
4	James Monroe	Apr 28, 1758	Westmoreland Co., Va.	July 4, 1831	New York, New York
5	John Quincy Adams	July 11, 1767	Quincy, Mass.	Feb 23, 1848	Washington, D.C.
6	Andrew Jackson	Mar 15, 1767	Waxhaws, No/So Carolina	June 8, 1845	Nashville, Tennessee
7	Martin Van Buren	Dec 5, 1782	Kinderhook, New York	July 24, 1862	Kinderhook, New York
8	William Henry Harrison	Feb 9, 1773	Charles City Co., Va.	Apr 4, 1841	Washington, D.C.
9	John Tyler	Mar 29, 1790	Charles City Co., Va.	Jan 18, 1862	Richmond, Va.
10	James K. Polk	Nov 2, 1795	Mecklenburg Co., N.C.	June 15, 1849	Nashville, Tennessee
11	Zachary Taylor	Nov 24, 1784	Orange County, Va.	July 9, 1850	Washington, D.C
12	Millard Fillmore	Jan 7, 1800	Cayuga Co., New York	Mar 8, 1874	Buffalo, New York
13	Franklin Pierce	Nov 23, 1804	Hillsborough, N.H.	Oct 8, 1869	Concord, New Hamp.
14	James Buchanan	Apr 23, 1791	Cove Gap, Pa.	June 1, 1868	Lancaster, Pa.
15	Abraham Lincoln	Feb 12, 1809	LaRue Co., Kentucky	Apr 15, 1865	Washington, D.C.

	PRESIDENT	BIRTH DATE	BIRTH PLACE	DEATH DATE	LOCATION OF DEATH
16	Andrew Johnson	Dec 29, 1808	Raleigh, North Carolina	July 31, 1875	Elizabethton, Tenn.
17	Ulysses S. Grant	Apr 27, 1822	Point Pleasant, Ohio	July 23, 1885	Wilton, New York
18	Rutherford B. Hayes	Oct 4, 1822	Delaware, Ohio	Jan 17, 1893	Fremont, Ohio
19	James A. Garfield	Nov 19, 1831	Cuyahoga Co., Ohio	Sep 19, 1881	Elberon, New Jersey
20	Chester Arthur	Oct 5, 1829	Fairfield, Vermont	Nov 18, 1886	New York, New York
21	Grover Cleveland	Mar 18, 1837	Caldwell, New Jersey	June 24, 1908	Princeton, New Jersey
22	Benjamin Harrison	Aug 20, 1833	North Bend, Ohio	Mar 13, 1901	Indianapolis, Indiana
23	William McKinley	Jan 29, 1843	Niles, Ohio	Sep 14, 1901	Buffalo, New York
24	Theodore Roosevelt	Oct 27, 1858	New York, New York	Jan 6, 1919	Oyster Bay, New York
25	William Howard Taft	Sep 15, 1857	Cincinnati, Ohio	Mar 8, 1930	Washington, D.C.
26	Woodrow Wilson	Dec 28, 1856	Staunton, Virginia	Feb 3, 1924	Washington, D.C.
27	Warren G. Harding	Nov 2, 1865	Morrow County, Ohio	Aug 2, 1923	San Francisco, Cal.
28	Calvin Coolidge	July 4, 1872	Plymouth, Vermont	Jan 5, 1933	Northampton, Mass.
29	Herbert Hoover	Aug 10, 1874	West Branch, Iowa	Oct 20, 1964	New York, New York
30	Franklin Roosevelt	Jan 30, 1882	Hyde Park, New York	Apr 12, 1945	Warm Springs, Ga.
31	Harry S. Truman	May 8, 1884	Lamar, Missouri	Dec 26, 1972	Kansas City, Missouri
32	Dwight Eisenhower	Oct 14, 1890	Denison, Texas	Mar 28, 1969	Washington, D.C.
33	John F. Kennedy	May 29, 1917	Brookline, Mass.	Nov 22, 1963	Dallas, Texas
34	Lyndon B. Johnson	Aug 27, 1908	Gillespie Co., Texas	Jan 22, 1973	Gillespie Co., Texas
35	Richard Nixon	Jan 9, 1913	Yorba Linda, Cal.	Apr 22, 1994	New York, New York
36	Gerald Ford	July 14, 1913	Omaha, Nebraska	Dec 26, 2006	Rancho Mirage, Cal.
37	Jimmy Carter	Oct 1, 1924	Plains, Georgia	NaN	NaN
38	Ronald Reagan	Feb 6, 1911	Tampico, Illinois	June 5, 2004	Los Angeles, Cal.
39	George Bush	June 12, 1924	Milton, Mass.	Nov 30, 2018	Houston, Texas
40	Bill Clinton	Aug 19, 1946	Hope, Arkansas	NaN	NaN
41	George W. Bush	July 6, 1946	New Haven, Conn.	NaN	NaN

	PRESIDENT	BIRTH DATE	BIRTH PLACE	DEATH DATE	LOCATION OF DEATH
42	Barack Obama	Aug 4, 1961	Honolulu, Hawaii	NaN	NaN
43	Donald Trump	June 14, 1946	New York, New York	NaN	NaN
44	Joe Biden	Nov 20, 1942	Scranton, Pa.	NaN	NaN
45	Reference:\nBirthplace, Birthdate, and Death i...	NaN	NaN	NaN	NaN

In [3]:
```python
#Remove unnecessary data and check
df.drop([len(df)-1], inplace=True)
df.tail()
```

Out[3]:

	PRESIDENT	BIRTH DATE	BIRTH PLACE	DEATH DATE	LOCATION OF DEATH
40	Bill Clinton	Aug 19, 1946	Hope, Arkansas	NaN	NaN
41	George W. Bush	July 6, 1946	New Haven, Conn.	NaN	NaN
42	Barack Obama	Aug 4, 1961	Honolulu, Hawaii	NaN	NaN
43	Donald Trump	June 14, 1946	New York, New York	NaN	NaN
44	Joe Biden	Nov 20, 1942	Scranton, Pa.	NaN	NaN

In [4]:
```python
#Add more rows to the dataframe
df = df.reindex(df.columns.tolist() + ['year_of_birth','lived_years', 'lived_months', 'lived_days'], axis=1)

#Change the date to datetime type (Year-Month—DAY) in order to calculate.
df['BIRTH DATE'] = pd.to_datetime(df['BIRTH DATE'])
df['DEATH DATE'] = pd.to_datetime(df['DEATH DATE'])

#Fill in blank column [year_of_birth]
df['year_of_birth'] = df['BIRTH DATE'].dt.year.astype(int)

#Fill in blank column [lived_years]
df['lived_years'] = df['DEATH DATE'] - df['BIRTH DATE']
df['lived_years'] = (df['lived_years'].fillna(pd.to_datetime(datetime.date(datetime.now())) - df['BIRTH DATE'])/ np.timedelta64(1, 'Y')).a

#Fill in blank column [lived_months]
df['lived_months'] = df['DEATH DATE'] - df['BIRTH DATE']
df['lived_months'] = (df['lived_months'].fillna(pd.to_datetime(datetime.date(datetime.now())) - df['BIRTH DATE'])/ np.timedelta64(1, 'M'))

#Fill in blank column [lived_days]
df['lived_days'] = df['DEATH DATE'] - df['BIRTH DATE']
df['lived_days'] = (df['lived_days'].fillna(pd.to_datetime(datetime.date(datetime.now())) - df['BIRTH DATE'])/ np.timedelta64(1, 'D')).ast

df
```

Out[4]:

	PRESIDENT	BIRTH DATE	BIRTH PLACE	DEATH DATE	LOCATION OF DEATH	year_of_birth	lived_years	lived_months	lived_days
0	George Washington	1732-02-22	Westmoreland Co., Va.	1799-12-14	Mount Vernon, Va.	1732	67	813	24767
1	John Adams	1735-10-30	Quincy, Mass.	1826-07-04	Quincy, Mass.	1735	90	1088	33119
2	Thomas Jefferson	1743-04-13	Albemarle Co., Va.	1826-07-04	Albemarle Co., Va.	1743	83	998	30397
3	James Madison	1751-03-16	Port Conway, Va.	1836-06-28	Orange Co., Va.	1751	85	1023	31150
4	James Monroe	1758-04-28	Westmoreland Co., Va.	1831-07-04	New York, New York	1758	73	878	26729
5	John Quincy Adams	1767-07-11	Quincy, Mass.	1848-02-23	Washington, D.C.	1767	80	967	29446
6	Andrew Jackson	1767-03-15	Waxhaws, No/So Carolina	1845-06-08	Nashville, Tennessee	1767	78	938	28574
7	Martin Van Buren	1782-12-05	Kinderhook, New York	1862-07-24	Kinderhook, New York	1782	79	955	29085
8	William Henry Harrison	1773-02-09	Charles City Co., Va.	1841-04-04	Washington, D.C.	1773	68	817	24890
9	John Tyler	1790-03-29	Charles City Co., Va.	1862-01-18	Richmond, Va.	1790	71	861	26227

	PRESIDENT	BIRTH DATE	BIRTH PLACE	DEATH DATE	LOCATION OF DEATH	year_of_birth	lived_years	lived_months	lived_days
10	James K. Polk	1795-11-02	Mecklenburg Co., N.C.	1849-06-15	Nashville, Tennessee	1795	53	643	19583
11	Zachary Taylor	1784-11-24	Orange County, Va.	1850-07-09	Washington, D.C	1784	65	787	23967
12	Millard Fillmore	1800-01-07	Cayuga Co., New York	1874-03-08	Buffalo, New York	1800	74	889	27088
13	Franklin Pierce	1804-11-23	Hillsborough, N.H.	1869-10-08	Concord, New Hamp.	1804	64	778	23695
14	James Buchanan	1791-04-23	Cove Gap, Pa.	1868-06-01	Lancaster, Pa.	1791	77	925	28163
15	Abraham Lincoln	1809-02-12	LaRue Co., Kentucky	1865-04-15	Washington, D.C.	1809	56	674	20516
16	Andrew Johnson	1808-12-29	Raleigh, North Carolina	1875-07-31	Elizabethton, Tenn.	1808	66	799	24320
17	Ulysses S. Grant	1822-04-27	Point Pleasant, Ohio	1885-07-23	Wilton, New York	1822	63	758	23098
18	Rutherford B. Hayes	1822-10-04	Delaware, Ohio	1893-01-17	Fremont, Ohio	1822	70	843	25673
19	James A. Garfield	1831-11-19	Cuyahoga Co., Ohio	1881-09-19	Elberon, New Jersey	1831	49	598	18202
20	Chester Arthur	1829-10-05	Fairfield, Vermont	1886-11-18	New York, New York	1829	57	685	20863
21	Grover Cleveland	1837-03-18	Caldwell, New Jersey	1908-06-24	Princeton, New Jersey	1837	71	855	26030
22	Benjamin Harrison	1833-08-20	North Bend, Ohio	1901-03-13	Indianapolis, Indiana	1833	67	810	24676
23	William McKinley	1843-01-29	Niles, Ohio	1901-09-14	Buffalo, New York	1843	58	703	21412
24	Theodore Roosevelt	1858-10-27	New York, New York	1919-01-06	Oyster Bay, New York	1858	60	722	21985
25	William Howard Taft	1857-09-15	Cincinnati, Ohio	1930-03-08	Washington, D.C.	1857	72	869	26471
26	Woodrow Wilson	1856-12-28	Staunton, Virginia	1924-02-03	Washington, D.C.	1856	67	805	24507
27	Warren G. Harding	1865-11-02	Morrow County, Ohio	1923-08-02	San Francisco, Cal.	1865	57	692	21091
28	Calvin Coolidge	1872-07-04	Plymouth, Vermont	1933-01-05	Northampton, Mass.	1872	60	726	22099
29	Herbert Hoover	1874-08-10	West Branch, Iowa	1964-10-20	New York, New York	1874	90	1082	32943
30	Franklin Roosevelt	1882-01-30	Hyde Park, New York	1945-04-12	Warm Springs, Ga.	1882	63	758	23082
31	Harry S. Truman	1884-05-08	Lamar, Missouri	1972-12-26	Kansas City, Missouri	1884	88	1063	32373
32	Dwight Eisenhower	1890-10-14	Denison, Texas	1969-03-28	Washington, D.C.	1890	78	941	28654
33	John F. Kennedy	1917-05-29	Brookline, Mass.	1963-11-22	Dallas, Texas	1917	46	557	16978
34	Lyndon B. Johnson	1908-08-27	Gillespie Co., Texas	1973-01-22	Gillespie Co., Texas	1908	64	772	23524
35	Richard Nixon	1913-01-09	Yorba Linda, Cal.	1994-04-22	New York, New York	1913	81	975	29688

	PRESIDENT	BIRTH DATE	BIRTH PLACE	DEATH DATE	LOCATION OF DEATH	year_of_birth	lived_years	lived_months	lived_days
36	Gerald Ford	1913-07-14	Omaha, Nebraska	2006-12-26	Rancho Mirage, Cal.	1913	93	1121	34133
37	Jimmy Carter	1924-10-01	Plains, Georgia	NaT	NaN	1924	98	1178	35855
38	Ronald Reagan	1911-02-06	Tampico, Illinois	2004-06-05	Los Angeles, Cal.	1911	93	1119	34088
39	George Bush	1924-06-12	Milton, Mass.	2018-11-30	Houston, Texas	1924	94	1133	34504
40	Bill Clinton	1946-08-19	Hope, Arkansas	NaT	NaN	1946	76	915	27863
41	George W. Bush	1946-07-06	New Haven, Conn.	NaT	NaN	1946	76	916	27907
42	Barack Obama	1961-08-04	Honolulu, Hawaii	NaT	NaN	1961	61	735	22399
43	Donald Trump	1946-06-14	New York, New York	NaT	NaN	1946	76	917	27929
44	Joe Biden	1942-11-20	Scranton, Pa.	NaT	NaN	1942	80	960	29231

Note:

In order to calculate the data that represent life span, I extract today's date and used for still alive presidents' lived days, months, and years.

In [5]:
```python
#Draw table for Top10 longest lived president
longest_lived = df.sort_values(by = 'lived_days', ascending=False, ignore_index=True)
longest_lived_table = (longest_lived.loc[0:9, ['PRESIDENT', 'lived_days']])
longest_lived_table.index = longest_lived_table.index+1
longest_lived_table
```

Out[5]:

	PRESIDENT	lived_days
1	Jimmy Carter	35855
2	George Bush	34504
3	Gerald Ford	34133
4	Ronald Reagan	34088
5	John Adams	33119
6	Herbert Hoover	32943
7	Harry S. Truman	32373
8	James Madison	31150
9	Thomas Jefferson	30397
10	Richard Nixon	29688

In [6]:
```python
#Draw table for Top10 shortest lived president
#Remove still alive presidents' data for better clearfication as [shortest_lived]
shortest_lived = df.dropna(axis=0, how='any', inplace=False)
shortest_lived = shortest_lived.sort_values(by = 'lived_days', ascending=True, ignore_index=True)
shortest_lived_table = (shortest_lived.loc[0:9, ['PRESIDENT', 'lived_days']])
shortest_lived_table.index = shortest_lived_table.index+1
shortest_lived_table
```

Out[6]:

	PRESIDENT	lived_days
1	John F. Kennedy	16978
2	James A. Garfield	18202
3	James K. Polk	19583
4	Abraham Lincoln	20516
5	Chester Arthur	20863
6	Warren G. Harding	21091
7	William McKinley	21412
8	Theodore Roosevelt	21985
9	Calvin Coolidge	22099
10	Franklin Roosevelt	23082

In [7]:
```python
#Statistic table for [lived days] and [lived years]
a = shortest_lived['lived_days']
b = shortest_lived['lived_years']
stat = {'Mean:':[a.mean(), b.mean()],
        #Error associated with each measurement(lived time) is the same(0)
        'Weighted average':[a.mean(), b.mean()],
        'Median':[a.median(), b.median()],
        'Mode':[(a//100).mode()*100,b.mode()],
        'Max':[a.max(), b.max()],
        'Min':[a.min(), b.min()],
        'Standard Deviation':[a.std(), b.std()]}
stat_table = pd.DataFrame(stat, index=['Lived Days', 'Lived Years']).astype(int)
stat_table.T
```

Out[7]:

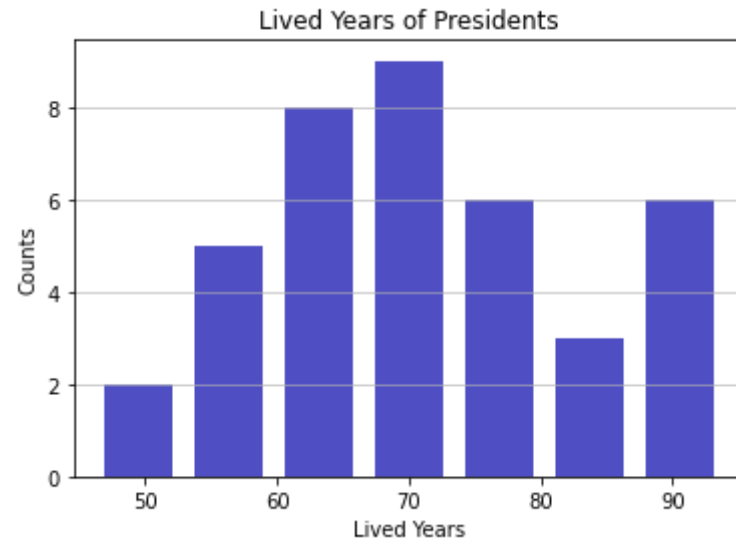|	Lived Days	Lived Years
Mean:	26097	71
Weighted average	26097	71
Median	25673	70
Mode	23000	67
Max	34504	94
Min	16978	46
Standard Deviation	4596	12

Note:

In order to better deliver the statistic

1. I calculate for both lived days and lived years which fits common sense.
2. Since we have only one sample of the data, the weighted average will stay the same as mean.
3. I round lived days to hundreds to calculate mode for lived_days. (For example: 24767 days round to 24000 days)

In [8]:
```python
#Plot the distribution
x = shortest_lived['lived_years']
fig, ax = plt.subplots()

#plot the histogram
x.plot.hist(density=False, ax=ax, bins=7, color = '#0504aa', alpha=0.7, rwidth=0.75, title='Lived Years of Presidents')
ax.set_xlabel('Lived Years')
ax.set_ylabel('Counts')
ax.grid(axis='y', alpha=0.75)
```


Note:

In order to better demonstrate the distribution, finer data(lived days) will show better details. However, lived years will fit better common sense. As a compromise, I first plot the histogram of lived days(h1) and than choosed the bin size that make the histogram plot of lived years(h2) looks the most like h1. Above plot is a visulization for h2.

Conclusion:

Distribution of Lived Days of Presidents shows strong pattern of normal distribution but with obvious increasing number of long-lived presidents. The longest lived president is Jimmy Carter with a life of 35855 days and the shortest lived president is John F. Kennedy who was assassinated with a life of 16978 days. Presidents had a avergae age of 71 years old. The standard deviation is 12 years or 4596 days means that 68% of presidents had a longevity between 59 to 83.

Future research topics:

1. The trend between birth year and lived days of presidents,
2. The lived days of presidents compares to lived days of U.S. male population

From these reseaches we could find more information regarding the factors may effect the lived days of presidents and population.

In [ ]: