

project: <https://github.com/clarkfitzg/sta141c-winter19/blob/master/project/project.md> (<https://github.com/clarkfitzg/sta141c-winter19/blob/master/project/project.md>)

```
In [13]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [5]: # difference between agg and agg_awards:
# agg: sum of total_obligation
# agg_awards: annual spending from awards table (more accurate)

agg = pd.read_csv("agg.csv")
agg_awards = pd.read_csv("agg_awards.csv")

# get rid of entries without fiscal year
agg_awards = agg_awards[agg_awards["fiscal_year"] > 0]

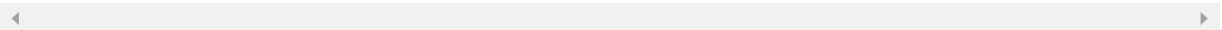
agg_awards
```

```
Out[5]:
```

	fiscal_year	awarding_agency_id	funding_agency_id	annual_spending
1445	1900.0	760.0	NaN	-1.130235e+08
1446	1973.0	930.0	NaN	9.180900e+04
1447	1992.0	1188.0	1188.0	4.580000e+05
1448	1993.0	1188.0	1188.0	3.920000e+05
1449	1994.0	1188.0	1188.0	4.610000e+05
1450	1995.0	1188.0	1188.0	4.060000e+05
1451	1996.0	48.0	48.0	3.080000e+05
1452	1996.0	1174.0	1174.0	3.000000e+04
1453	1996.0	1188.0	1188.0	9.400000e+04
1454	1997.0	633.0	633.0	1.750000e+05
1455	1997.0	1174.0	1174.0	7.248330e+05
1456	1997.0	1188.0	1188.0	8.781560e+05
1457	1997.0	1223.0	1223.0	2.474510e+05
1458	1998.0	90.0	90.0	8.000000e+04
1459	1998.0	633.0	633.0	3.220000e+06
1460	1998.0	862.0	862.0	3.000000e+03
1461	1998.0	1174.0	1174.0	2.367801e+06
1462	1998.0	1188.0	1188.0	5.567760e+05
1463	1998.0	1196.0	1196.0	-2.700980e+05
1464	1998.0	1217.0	1217.0	-3.669400e+04
1465	1998.0	1235.0	1235.0	-6.080000e+04
1466	1999.0	315.0	NaN	2.800000e+04
1467	1999.0	633.0	633.0	4.028006e+06
1468	1999.0	635.0	635.0	2.635077e+06
1469	1999.0	636.0	636.0	2.500000e+04
1470	1999.0	801.0	801.0	2.866502e+05
1471	1999.0	1174.0	1174.0	1.686248e+07
1472	1999.0	1188.0	1188.0	1.296843e+06
1473	1999.0	1196.0	1196.0	5.557890e+05
1474	1999.0	1217.0	1217.0	-1.603230e+05
...	...	...	...	...

20818	2018.0	1223.0	1206.0	7.150454e+05
20819	2018.0	1223.0	1217.0	6.163005e+06
20820	2018.0	1223.0	1219.0	3.117951e+07
20821	2018.0	1223.0	1223.0	3.269012e+07
20822	2018.0	1223.0	1225.0	5.806130e+06
20823	2018.0	1223.0	1231.0	1.493884e+07
20824	2018.0	1225.0	1173.0	2.616576e+05
20825	2018.0	1225.0	1225.0	6.911198e+10
20826	2018.0	1227.0	1173.0	1.841051e+06
20827	2018.0	1227.0	1174.0	5.585812e+04
20828	2018.0	1227.0	1196.0	3.797564e+06
20829	2018.0	1227.0	1227.0	3.104921e+07
20830	2018.0	1228.0	1226.0	9.769645e+05
20831	2018.0	1228.0	1228.0	2.659156e+07
20832	2018.0	1231.0	NaN	1.227386e+08
20833	2018.0	1231.0	1173.0	4.618677e+08
20834	2018.0	1231.0	1188.0	2.067912e+05
20835	2018.0	1231.0	1217.0	3.471740e+07
20836	2018.0	1231.0	1225.0	4.622344e+06
20837	2018.0	1231.0	1231.0	1.843594e+08
20838	2018.0	1231.0	1232.0	1.212659e+07
20839	2018.0	1233.0	NaN	4.238133e+07
20840	2018.0	1233.0	1233.0	2.424427e+06
20841	2018.0	1234.0	1188.0	4.557505e+08
20842	2018.0	1234.0	1234.0	4.323471e+10
20843	2018.0	1235.0	1173.0	1.598721e+05
20844	2018.0	1235.0	1174.0	2.925172e+07
20845	2018.0	1235.0	1188.0	3.554958e+07
20846	2018.0	1235.0	1235.0	1.893940e+09
20847	2018.0	1423.0	NaN	1.320119e+07

19403 rows x 4 columns



In [11]: *# created parties.xlsx based on [https://en.wikipedia.org/wiki/Party\\_divisions\\_of\\_United\\_States\\_Congresses](https://en.wikipedia.org/wiki/Party_divisions_of_United_States_Congresses)*

```
parties = pd.read_excel("parties.xlsx")
parties.columns.values[[0]] = "fiscal_year"
parties
```

Out[11]:

	fiscal_year	house	senate	president
0	2000	R	R	D
1	2001	R	R	R
2	2002	R	R	R
3	2003	R	R	R
4	2004	R	R	R
5	2005	R	R	R
6	2006	R	R	R
7	2007	D	D	R
8	2008	D	D	R
9	2009	D	D	D

#	year	house	senate	president
10	2010	D	D	D
11	2011	R	D	D
12	2012	R	D	D
13	2013	R	D	D
14	2014	R	D	D
15	2015	R	R	D
16	2016	R	R	D
17	2017	R	R	R
18	2018	R	R	R
19	2019	D	R	R

```
In [16]: # combine party data and award data
data = pd.merge(agg_awards, parties, on = "fiscal_year")
data
```

Out[16]:

	fiscal_year	awarding_agency_id	funding_agency_id	annual_spending	house	senate	president
0	2000.0	48.0	48.0	1.227600e+07	R	R	D
1	2000.0	67.0	NaN	2.281240e+05	R	R	D
2	2000.0	89.0	89.0	6.352000e+06	R	R	D
3	2000.0	90.0	90.0	3.811000e+06	R	R	D
4	2000.0	99.0	99.0	8.140000e+05	R	R	D
5	2000.0	100.0	100.0	4.354970e+05	R	R	D
6	2000.0	104.0	104.0	7.267100e+07	R	R	D
7	2000.0	105.0	105.0	1.217500e+07	R	R	D
8	2000.0	106.0	106.0	1.767849e+09	R	R	D
9	2000.0	110.0	95.0	2.230000e+05	R	R	D
10	2000.0	110.0	110.0	5.181861e+08	R	R	D
11	2000.0	111.0	111.0	5.130000e+07	R	R	D
12	2000.0	114.0	114.0	7.187900e+07	R	R	D
13	2000.0	116.0	116.0	2.675100e+07	R	R	D
14	2000.0	118.0	118.0	7.905830e+08	R	R	D
15	2000.0	119.0	119.0	3.869800e+07	R	R	D
16	2000.0	120.0	120.0	2.015800e+07	R	R	D
17	2000.0	121.0	121.0	1.762500e+07	R	R	D
18	2000.0	133.0	NaN	2.901912e+04	R	R	D
19	2000.0	133.0	133.0	3.592756e+06	R	R	D
20	2000.0	149.0	NaN	1.869011e+06	R	R	D
21	2000.0	149.0	149.0	5.926130e+07	R	R	D
22	2000.0	150.0	150.0	6.471352e+06	R	R	D
23	2000.0	153.0	153.0	6.782696e+06	R	R	D
24	2000.0	158.0	158.0	1.003323e+06	R	R	D
25	2000.0	163.0	163.0	7.348200e+04	R	R	D
26	2000.0	168.0	168.0	3.297580e+06	R	R	D
27	2000.0	175.0	NaN	1.376800e+05	R	R	D
28	2000.0	175.0	175.0	1.290005e+07	R	R	D
29	2000.0	176.0	176.0	5.961340e+05	R	R	D
...	...	...	...	...	...	...	...

19342	2018.0	1223.0	1206.0	7.150454e+05	R	R	R
19343	2018.0	1223.0	1217.0	6.163005e+06	R	R	R
19344	2018.0	1223.0	1219.0	3.117951e+07	R	R	R
19345	2018.0	1223.0	1223.0	3.269012e+07	R	R	R
19346	2018.0	1223.0	1225.0	5.806130e+06	R	R	R
19347	2018.0	1223.0	1231.0	1.493884e+07	R	R	R
19348	2018.0	1225.0	1173.0	2.616576e+05	R	R	R
19349	2018.0	1225.0	1225.0	6.911198e+10	R	R	R
19350	2018.0	1227.0	1173.0	1.841051e+06	R	R	R
19351	2018.0	1227.0	1174.0	5.585812e+04	R	R	R
19352	2018.0	1227.0	1196.0	3.797564e+06	R	R	R
19353	2018.0	1227.0	1227.0	3.104921e+07	R	R	R
19354	2018.0	1228.0	1226.0	9.769645e+05	R	R	R
19355	2018.0	1228.0	1228.0	2.659156e+07	R	R	R
19356	2018.0	1231.0	NaN	1.227386e+08	R	R	R
19357	2018.0	1231.0	1173.0	4.618677e+08	R	R	R
19358	2018.0	1231.0	1188.0	2.067912e+05	R	R	R
19359	2018.0	1231.0	1217.0	3.471740e+07	R	R	R
19360	2018.0	1231.0	1225.0	4.622344e+06	R	R	R
19361	2018.0	1231.0	1231.0	1.843594e+08	R	R	R
19362	2018.0	1231.0	1232.0	1.212659e+07	R	R	R
19363	2018.0	1233.0	NaN	4.238133e+07	R	R	R
19364	2018.0	1233.0	1233.0	2.424427e+06	R	R	R
19365	2018.0	1234.0	1188.0	4.557505e+08	R	R	R
19366	2018.0	1234.0	1234.0	4.323471e+10	R	R	R
19367	2018.0	1235.0	1173.0	1.598721e+05	R	R	R
19368	2018.0	1235.0	1174.0	2.925172e+07	R	R	R
19369	2018.0	1235.0	1188.0	3.554958e+07	R	R	R
19370	2018.0	1235.0	1235.0	1.893940e+09	R	R	R
19371	2018.0	1423.0	NaN	1.320119e+07	R	R	R

19372 rows x 7 columns

```
In [49]: yearly_spending = data.groupby("fiscal_year").aggregate({"annual_spending" : "sum"})
yearly_spending = yearly_spending.reset_index()

# get rid of 2000 and 2018 because the data is incomplete
yearly_spending_full = yearly_spending
yearly_spending_full = yearly_spending_full[yearly_spending_full["fiscal_year"] != 2000]
yearly_spending_full = yearly_spending_full[yearly_spending_full["fiscal_year"] != 2018]
yearly_spending_full
```

Out[49]:

	fiscal_year	annual_spending
1	2001.0	8.945262e+10
2	2002.0	1.280627e+11
3	2003.0	1.788192e+11
4	2004.0	1.595543e+11
5	2005.0	1.981982e+11
6	2006.0	2.068323e+11
7	2007.0	9.786393e+11
8	2008.0	1.377017e+12

9	2009.0	2.605274e+12
10	2010.0	2.161247e+12
11	2011.0	3.104149e+12
12	2012.0	3.655370e+12
13	2013.0	3.836789e+12
14	2014.0	2.728217e+12
15	2015.0	2.941107e+12
16	2016.0	3.180447e+12
17	2017.0	4.153457e+12

```
In [50]: ax = sns.lineplot("fiscal_year", "annual_spending", data = yearly_spending_full)
ax.set(title = "Total government spending from 2001 to 2017",
      xlabel = "Year",
      ylabel = "Total annual spending (trillions)")
```

```
Out[50]: [Text(0, 0.5, 'Total annual spending (trillions)'),
Text(0.5, 0, 'Year'),
Text(0.5, 1.0, 'Total government spending from 2001 to 2017')]
```

