

IST 387 HW 5

Copyright 2021, Jeffrey Stanton, Jeffrey Saltz, and Jasmina Tacheva

```
In [1]: # Enter your name here: Ezra Cohen
```

Attribution statement: (choose only one and delete the rest)

```
In [2]: # 2. I did this homework with help from the book and the professor and these Internet sources:google to describe col
```

(Chapter 11 of Introduction to Data Science)

Reminders of things to practice from previous weeks:

Descriptive statistics: mean() max() min()

Coerce to numeric: as.numeric()

Part 1: Use the Starter Code

Below, I have provided a starter file to help you.

Each of these lines of code **must be commented** (the comment must that explains what is going on, so that I know you understand the code and results).

```
In [10]: library(RCurl)
library(jsonlite)
#These 2 lines make sure that the packages necessary to run the code are able to be used
dataset <- getURL("https://ist387.s3.us-east-2.amazonaws.com/data/role.json")
#gets information/results from the url and stores it in a varriable
readlines <- jsonlite::fromJSON(dataset)
#converts the results to a dataframe in R
df <- readlines$objects$person
#sets df to the data in person
```

A. Explore the **df** dataframe (e.g., using View() or whatever you think is best).

In [14]:

```
#View(df) this doesn't work  
df #The only reason I knew to do this was because I've used jupyter notebooks before
```

A data.frame: 100 × 1

	bioguideid	birthday	cspanid	firstname	gender	gender_label	lastname	link
	<chr>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<chr>
1	C000880	1951-05-20	26440	Michael	male	Male	Crapo	https://www.govtrack.us/congress/members/michael_crapo/300030
2	G000386	1933-09-17	1167	Charles	male	Male	Grassley	https://www.govtrack.us/congress/members/charles_grassley/300048
3	L000174	1940-03-31	1552	Patrick	male	Male	Leahy	https://www.govtrack.us/congress/members/patrick_leahy/300065
4	M001153	1957-05-22	1004138	Lisa	female	Female	Murkowski	https://www.govtrack.us/congress/members/lisa_murkowski/300075
5	M001111	1950-10-11	25277	Patty	female	Female	Murray	https://www.govtrack.us/congress/members/patty_murray/300076
6	S000148	1950-11-23	5929	Charles	male	Male	Schumer	https://www.govtrack.us/congress/members/charles_schumer/300087
7	S000320	1934-05-06	1859	Richard	male	Male	Shelby	https://www.govtrack.us/congress/members/richard_shelby/300089
8	W000779	1949-05-03	1962	Ron	male	Male	Wyden	https://www.govtrack.us/congress/members/ron_wyden/300100
9	B000575	1950-01-10	45465	Roy	male	Male	Blunt	https://www.govtrack.us/congress/members/roy_blunt/400034
10	B001236	1950-12-10	92069	John	male	Male	Boozman	https://www.govtrack.us/congress/members/john_boozman/400040
11	B001135	1955-11-30	31054	Richard	male	Male	Burr	https://www.govtrack.us/congress/members/richard_burr/400054

	bioguideid	birthday	cspanid	firstname	gender	gender_label	lastname	link
	<chr>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<chr>
12	M000934	1954-05-29	45469	Jerry	male	Male	Moran	https://www.govtrack.us/congress/members/jerry_moran/400284
13	P000449	1955-12-19	31819	Robert	male	Male	Portman	https://www.govtrack.us/congress/members/robert_portman/400325
14	T000461	1961-11-17	7958	Patrick	male	Male	Toomey	https://www.govtrack.us/congress/members/patrick_toomey/400408
15	V000128	1959-01-10	20756	Chris	male	Male	Van Hollen	https://www.govtrack.us/congress/members/chris_van_hollen/400415
16	T000250	1961-01-07	45552	John	male	Male	Thune	https://www.govtrack.us/congress/members/john_thune/400546
17	B001267	1964-11-28	1031622	Michael	male	Male	Bennet	https://www.govtrack.us/congress/members/michael_bennet/412330
18	Y000064	1972-08-24	1033743	Todd	male	Male	Young	https://www.govtrack.us/congress/members/todd_young/412428
19	L000575	1968-03-04	1033847	James	male	Male	Lankford	https://www.govtrack.us/congress/members/james_lankford/412464
20	S001184	1965-09-19	623506	Tim	male	Male	Scott	https://www.govtrack.us/congress/members/tim_scott/412471
21	B001277	1946-02-13	21799	Richard	male	Male	Blumenthal	https://www.govtrack.us/congress/members/richard_blumenthal/412490
22	R000595	1971-05-28	87599	Marco	male	Male	Rubio	https://www.govtrack.us/congress/members/marco_rubio/412491
23	P000603	1963-01-07	9265241	Rand	male	Male	Paul	https://www.govtrack.us/congress/members/rand_paul/412492
24	H001061	1957-03-13	85233	John	male	Male	Hoeven	https://www.govtrack.us/congress/members/john_hoeven/412494
25	L000577	1971-06-04	9267977	Mike	male	Male	Lee	https://www.govtrack.us/congress/members/mike_lee/412495

	bioguideid	birthday	cspanid	firstname	gender	gender_label	lastname	link
	<chr>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<chr>
26	J000293	1955-04-08	62835	Ron	male	Male	Johnson	https://www.govtrack.us/congress/members/ron_johnson/412496
27	S001194	1972-10-20	87784	Brian	male	Male	Schatz	https://www.govtrack.us/congress/members/brian_schatz/412507
28	D000622	1968-03-12	94484	Tammy	female	Female	Duckworth	https://www.govtrack.us/congress/members/tammy_duckworth/412533
29	K000393	1951-11-21	1011723	John	male	Male	Kennedy	https://www.govtrack.us/congress/members/john_kennedy/412679
30	H001076	1958-02-27	67481	Margaret	female	Female	Hassan	https://www.govtrack.us/congress/members/margaret_hassan/412680
:	:	:	:	:	:	:	:	:
71	M000355	1942-02-20	2351	Mitch	male	Male	McConnell	https://www.govtrack.us/congress/members/mitch_mcconnell/300072
72	M001176	1956-10-24	1029842	Jeff	male	Male	Merkley	https://www.govtrack.us/congress/members/jeff_merkley/412325
73	R000122	1949-11-12	24239	John	male	Male	Reed	https://www.govtrack.us/congress/members/john_reed/300081
74	R000584	1943-05-03	1020034	James	male	Male	Risch	https://www.govtrack.us/congress/members/james_risch/412322
75	S001181	1947-01-28	22850	Jeanne	female	Female	Shaheen	https://www.govtrack.us/congress/members/jeanne_shaheen/412323
76	W000805	1954-12-15	7630	Mark	male	Male	Warner	https://www.govtrack.us/congress/members/mark_warner/412321
77	C001088	1963-09-09	9269028	Christopher	male	Male	Coons	https://www.govtrack.us/congress/members/christopher_coons/412390
78	C001047	1953-11-26	83737	Shelley	female	Female	Capito	https://www.govtrack.us/congress/members/shelley_capito/400061
79	C001075	1957-09-28	1030546	Bill	male	Male	Cassidy	https://www.govtrack.us/congress/members/bill_cassidy/412269

	bioguideid	birthday	cspanid	firstname	gender	gender_label	lastname	link
	<chr>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<chr>
80	L000570	1972-06-07	1031351	Ben	male	Male	Lujan	https://www.govtrack.us/congress/members/ben_lujan/412293
81	M000133	1946-07-11	260	Edward	male	Male	Markey	https://www.govtrack.us/congress/members/edward_markey/400253
82	P000595	1958-12-01	50199	Gary	male	Male	Peters	https://www.govtrack.us/congress/members/gary_peters/412305
83	C001095	1977-05-13	63928	Tom	male	Male	Cotton	https://www.govtrack.us/congress/members/tom_cotton/412508
84	D000618	1962-08-20	1034037	Steve	male	Male	Daines	https://www.govtrack.us/congress/members/steve_daines/412549
85	B001288	1969-04-27	84679	Cory	male	Male	Booker	https://www.govtrack.us/congress/members/cory_booker/412598
86	S001198	1964-11-13	1023262	Dan	male	Male	Sullivan	https://www.govtrack.us/congress/members/dan_sullivan/412665
87	E000295	1970-07-01	75342	Joni	female	Female	Ernst	https://www.govtrack.us/congress/members/joni_ernst/412667
88	T000476	1960-08-30	77055	Thom	male	Male	Tillis	https://www.govtrack.us/congress/members/thom_tillis/412668
89	R000605	1954-10-24	78317	Mike	male	Male	Rounds	https://www.govtrack.us/congress/members/mike_rounds/412669
90	S001197	1972-02-22	77429	Benjamin	male	Male	Sasse	https://www.govtrack.us/congress/members/benjamin_sasse/412671
91	M001198	1960-08-09	103425	Roger	male	Male	Marshall	https://www.govtrack.us/congress/members/roger_marshall/412704
92	S001203	1958-03-04	111313	Tina	female	Female	Smith	https://www.govtrack.us/congress/members/tina_smith/412742
93	H001079	1959-05-10	113208	Cindy	female	Female	Hyde-Smith	https://www.govtrack.us/congress/members/cindy_hyde_smith/412743

	bioguideid	birthday	cspanid	firstname	gender	gender_label	lastname	link
	<chr>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<chr>
94	L000571	1954-09-10	1031365	Cynthia	female	Female	Lummis	https://www.govtrack.us/congress/members/cynthia_lummis/412294
95	T000278	1954-09-18	NA	Tommy	male	Male	Tuberville	https://www.govtrack.us/congress/members/tommy_tuberville/456796
96	H000273	1952-02-07	NA	John	male	Male	Hickenlooper	https://www.govtrack.us/congress/members/john_hickenlooper/456797
97	H000601	1959-08-14	NA	Bill	male	Male	Hagerty	https://www.govtrack.us/congress/members/bill_hagerty/456798
98	P000145	1973-03-22	NA	Alejandro	male	Male	Padilla	https://www.govtrack.us/congress/members/alejandro_padilla/456856
99	O000174	1987-02-16	NA	Jon	male	Male	Ossoff	https://www.govtrack.us/congress/members/jon_ossoff/456857
100	W000790	1969-07-23	NA	Raphael	male	Male	Warnock	https://www.govtrack.us/congress/members/raphael_warnock/456858

B. Explain the dataset

- o What is the dataset about?
- o How many rows are there and what does a row represent?
- o How many columns and what does each column represent?

This data set is about a hundred different senators with each row being for one of those senators, 100 rows, There are 17 different columns each one giving different information about the senator, bioguideid Is an id each member has chosen for themselves and each one is different, birthday is the year month and day they were born, cspanid Is an ID given to Veterans and military service members, Firstname is their first name, gender is their gender, gender_label is the gender they have chosen to label themselves as, lastname is their last name, Link gives a link to a page of information about them, Middlename gives the first initial of their middle name if they have one, name gives their full name the fact that they are a senator and a bit of other information, I don't know what namemod is it seems to be an empty column and looking it up on Google doesn't help, Nickname gives nicknames for anyone who has one, osid is one that I don't know what it is again because when I looked it up on Google it said it was about the intelligence of animals which I don't think it's something that we're going to label Senators with, pvsid again I don't know what it is

because when I look it up on Google it says it's a gaming ID which is again not information I think that we care about for senators, Sortname has all the information from name but in a more sensible order, TwitterID gives their account on Twitter, YouTubeID gives their account on YouTube

Part 2: Investigate the resulting dataframe

C. How many senators are women?

```
In [30]: womensenators<-df[c(which(df$gender=="female")),]  
#length(womensenators)#I think this is only saying number of columns #it is I'll tag it out now  
nrow(womensenators)  
womensenators
```

24

A data.frame: 24 × 1

	bioguideid	birthday	cspanid	firstname	gender	gender_label	lastname	link
	<chr>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<chr>
4	M001153	1957-05-22	1004138	Lisa	female	Female	Murkowski	https://www.govtrack.us/congress/members/lisa_murkowski/300075
5	M001111	1950-10-11	25277	Patty	female	Female	Murray	https://www.govtrack.us/congress/members/patty_murray/300076
28	D000622	1968-03-12	94484	Tammy	female	Female	Duckworth	https://www.govtrack.us/congress/members/tammy_duckworth/412533
30	H001076	1958-02-27	67481	Margaret	female	Female	Hassan	https://www.govtrack.us/congress/members/margaret_hassan/412680
31	C001113	1964-03-29	105698	Catherine	female	Female	Cortez Masto	https://www.govtrack.us/congress/members/catherine_cortez_masto/412681
32	C000127	1958-10-13	26137	Maria	female	Female	Cantwell	https://www.govtrack.us/congress/members/maria_cantwell/300018
34	F000062	1933-06-22	13061	Dianne	female	Female	Feinstein	https://www.govtrack.us/congress/members/dianne_feinstein/300043
35	S000770	1950-04-29	45451	Debbie	female	Female	Stabenow	https://www.govtrack.us/congress/members/debbie_stabenow/300093

	bioguideid	birthday	cspanid	firstname	gender	gender_label	lastname	link	r
	<chr>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<chr>	
36	B001230	1962-02-11	57884	Tammy	female	Female	Baldwin	https://www.govtrack.us/congress/members/tammy_baldwin/400013	
37	B001243	1952-06-06	31226	Marsha	female	Female	Blackburn	https://www.govtrack.us/congress/members/marsha_blackburn/400032	
44	H001042	1947-11-03	91216	Mazie	female	Female	Hirono	https://www.govtrack.us/congress/members/mazie_hirono/412200	
45	G000555	1966-12-09	1022862	Kirsten	female	Female	Gillibrand	https://www.govtrack.us/congress/members/kirsten_gillibrand/412223	
46	K000367	1960-05-25	83701	Amy	female	Female	Klobuchar	https://www.govtrack.us/congress/members/amy_klobuchar/412242	
53	S001191	1976-07-12	68489	Kyrsten	female	Female	Sinema	https://www.govtrack.us/congress/members/kyrsten_sinema/412509	
54	W000817	1949-06-22	1023023	Elizabeth	female	Female	Warren	https://www.govtrack.us/congress/members/elizabeth_warren/412542	
57	F000463	1951-03-01	1034067	Deb	female	Female	Fischer	https://www.govtrack.us/congress/members/deb_fischer/412556	
60	R000608	1957-08-02	104738	Jacky	female	Female	Rosen	https://www.govtrack.us/congress/members/jacky_rosen/412715	
66	C001035	1952-12-07	45738	Susan	female	Female	Collins	https://www.govtrack.us/congress/members/susan_collins/300025	
75	S001181	1947-01-28	22850	Jeanne	female	Female	Shaheen	https://www.govtrack.us/congress/members/jeanne_shaheen/412323	
78	C001047	1953-11-26	83737	Shelley	female	Female	Capito	https://www.govtrack.us/congress/members/shelley_capito/400061	
87	E000295	1970-07-01	75342	Joni	female	Female	Ernst	https://www.govtrack.us/congress/members/joni_ernst/412667	
92	S001203	1958-03-04	111313	Tina	female	Female	Smith	https://www.govtrack.us/congress/members/tina_smith/412742	
93	H001079	1959-05-10	113208	Cindy	female	Female	Hyde-Smith	https://www.govtrack.us/congress/members/cindy_hyde_smith/412743	

	bioguideid	birthday	cspanid	firstname	gender	gender_label	lastname	link	r
	<chr>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<chr>	
94	L000571	1954-09-10	1031365	Cynthia	female	Female	Lummis	https://www.govtrack.us/congress/members/cynthia_lummis/412294	

D. How many senators have a YouTube account?

```
In [28]: YTsenators<-is.na(df$youtubeid)
YTsenators<-df[c(which(YTsenators=="FALSE")),]
nrow(YTsenators)
```

73

E. How many women senators have a YouTube account?

```
In [33]: YTwomensenators<-YTsenators[-c(which(df$gender=="male")),]#When I just did it with the female the same way I did when
nrow(YTwomensenators)
```

18

A data.frame: 18 × 17

	bioguideid	birthday	cspanid	firstname	gender	gender_label	lastname	link	mi
	<chr>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<chr>	
4	M001153	1957-05-22	1004138	Lisa	female	Female	Murkowski	https://www.govtrack.us/congress/members/lisa_murkowski/300075	
5	M001111	1950-10-11	25277	Patty	female	Female	Murray	https://www.govtrack.us/congress/members/patty_murray/300076	
28	D000622	1968-03-12	94484	Tammy	female	Female	Duckworth	https://www.govtrack.us/congress/members/tammy_duckworth/412533	
33	C000174	1947-01-23	663	Thomas	male	Male	Carper	https://www.govtrack.us/congress/members/thomas_carper/300019	

	bioguideid	birthday	cspanid	firstname	gender	gender_label	lastname	link	mi
	<chr>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<chr>	
34	F000062	1933-06-22	13061	Dianne	female	Female	Feinstein	https://www.govtrack.us/congress/members/dianne_feinstein/300043	
35	S000770	1950-04-29	45451	Debbie	female	Female	Stabenow	https://www.govtrack.us/congress/members/debbie_stabenow/300093	
37	B001243	1952-06-06	31226	Marsha	female	Female	Blackburn	https://www.govtrack.us/congress/members/marsha_blackburn/400032	
38	B000944	1952-11-09	5051	Sherrod	male	Male	Brown	https://www.govtrack.us/congress/members/sherrod_brown/400050	
39	C000141	1943-10-05	4004	Benjamin	male	Male	Cardin	https://www.govtrack.us/congress/members/benjamin_cardin/400064	
40	M000639	1954-01-01	29608	Robert	male	Male	Menendez	https://www.govtrack.us/congress/members/robert_menendez/400272	
47	T000464	1956-08-21	1020176	Jon	male	Male	Tester	https://www.govtrack.us/congress/members/jon_tester/412244	
48	C001070	1960-04-13	47036	Robert	male	Male	Casey	https://www.govtrack.us/congress/members/robert_casey/412246	
49	W000802	1955-10-20	92235	Sheldon	male	Male	Whitehouse	https://www.govtrack.us/congress/members/sheldon_whitehouse/412247	
56	C001096	1961-01-21	7600	Kevin	male	Male	Cramer	https://www.govtrack.us/congress/members/kevin_cramer/412555	
57	F000463	1951-03-01	1034067	Deb	female	Female	Fischer	https://www.govtrack.us/congress/members/deb_fischer/412556	
66	C001035	1952-12-07	45738	Susan	female	Female	Collins	https://www.govtrack.us/congress/members/susan_collins/300025	

	bioguideid	birthday	cspanid	firstname	gender	gender_label	lastname	link	mi
	<chr>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<chr>	
69	G000359	1955-07-09	36782	Lindsey	male	Male	Graham	https://www.govtrack.us/congress/members/lindsey_graham/300047	
76	W000805	1954-12-15	7630	Mark	male	Male	Warner	https://www.govtrack.us/congress/members/mark_warner/412321	

F. Create a new dataframe called **youtubeWomen** that only includes women senators who have a YouTube account.

In [34]:

```
youtubeWomen<-data.frame(YTwomensenators)
youtubeWomen
```

A data.frame: 18 × 17

	bioguideid	birthday	cspanid	firstname	gender	gender_label	lastname	link	mi
	<chr>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<chr>	
4	M001153	1957-05-22	1004138	Lisa	female	Female	Murkowski	https://www.govtrack.us/congress/members/lisa_murkowski/300075	
5	M001111	1950-10-11	25277	Patty	female	Female	Murray	https://www.govtrack.us/congress/members/patty_murray/300076	
28	D000622	1968-03-12	94484	Tammy	female	Female	Duckworth	https://www.govtrack.us/congress/members/tammy_duckworth/412533	
33	C000174	1947-01-23	663	Thomas	male	Male	Carper	https://www.govtrack.us/congress/members/thomas_carper/300019	
34	F000062	1933-06-22	13061	Dianne	female	Female	Feinstein	https://www.govtrack.us/congress/members/dianne_feinstein/300043	
35	S000770	1950-04-29	45451	Debbie	female	Female	Stabenow	https://www.govtrack.us/congress/members/debbie_stabenow/300093	

	bioguideid	birthday	cspanid	firstname	gender	gender_label	lastname	link	mi
	<chr>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<chr>	
37	B001243	1952-06-06	31226	Marsha	female	Female	Blackburn	https://www.govtrack.us/congress/members/marsha_blackburn/400032	
38	B000944	1952-11-09	5051	Sherrod	male	Male	Brown	https://www.govtrack.us/congress/members/sherrod_brown/400050	
39	C000141	1943-10-05	4004	Benjamin	male	Male	Cardin	https://www.govtrack.us/congress/members/benjamin_cardin/400064	
40	M000639	1954-01-01	29608	Robert	male	Male	Menendez	https://www.govtrack.us/congress/members/robert_menendez/400272	
47	T000464	1956-08-21	1020176	Jon	male	Male	Tester	https://www.govtrack.us/congress/members/jon_tester/412244	
48	C001070	1960-04-13	47036	Robert	male	Male	Casey	https://www.govtrack.us/congress/members/robert_casey/412246	
49	W000802	1955-10-20	92235	Sheldon	male	Male	Whitehouse	https://www.govtrack.us/congress/members/sheldon_whitehouse/412247	
56	C001096	1961-01-21	7600	Kevin	male	Male	Cramer	https://www.govtrack.us/congress/members/kevin_cramer/412555	
57	F000463	1951-03-01	1034067	Deb	female	Female	Fischer	https://www.govtrack.us/congress/members/deb_fischer/412556	
66	C001035	1952-12-07	45738	Susan	female	Female	Collins	https://www.govtrack.us/congress/members/susan_collins/300025	
69	G000359	1955-07-09	36782	Lindsey	male	Male	Graham	https://www.govtrack.us/congress/members/lindsey_graham/300047	
76	W000805	1954-12-15	7630	Mark	male	Male	Warner	https://www.govtrack.us/congress/members/mark_warner/412321	

G. What does running this line of code do? Explain in a comment:

```
In [37]: youtubewomen$year <- substr(youtubewomen$birthday,1,4)
youtubewomen$year
#Just by looking at the results what I think is happening is that it is extracting from the first character to the fo
```

'1957' · '1950' · '1968' · '1947' · '1933' · '1950' · '1952' · '1952' · '1943' · '1954' · '1956' · '1960' · '1955' · '1961' · '1951' · '1952' · '1955' · '1954'

H. Use this new variable to calculate the mean **birthyear** in **youtubewomen**.

Hint: You may need to convert it to numeric first.

```
In [38]: mean(as.numeric(youtubewomen$year))
```

1952.77777777778

I. Make a histogram of the **birthyears** of senators in **youtubewomen**. Add a comment describing the shape of the distribution.

```
In [40]: hist(as.numeric(youtubewomen$year))
#The shape of the distribution is a bell-shaped or normal distribution with the center at 1950 to 1955(one bar is 5 y
```

