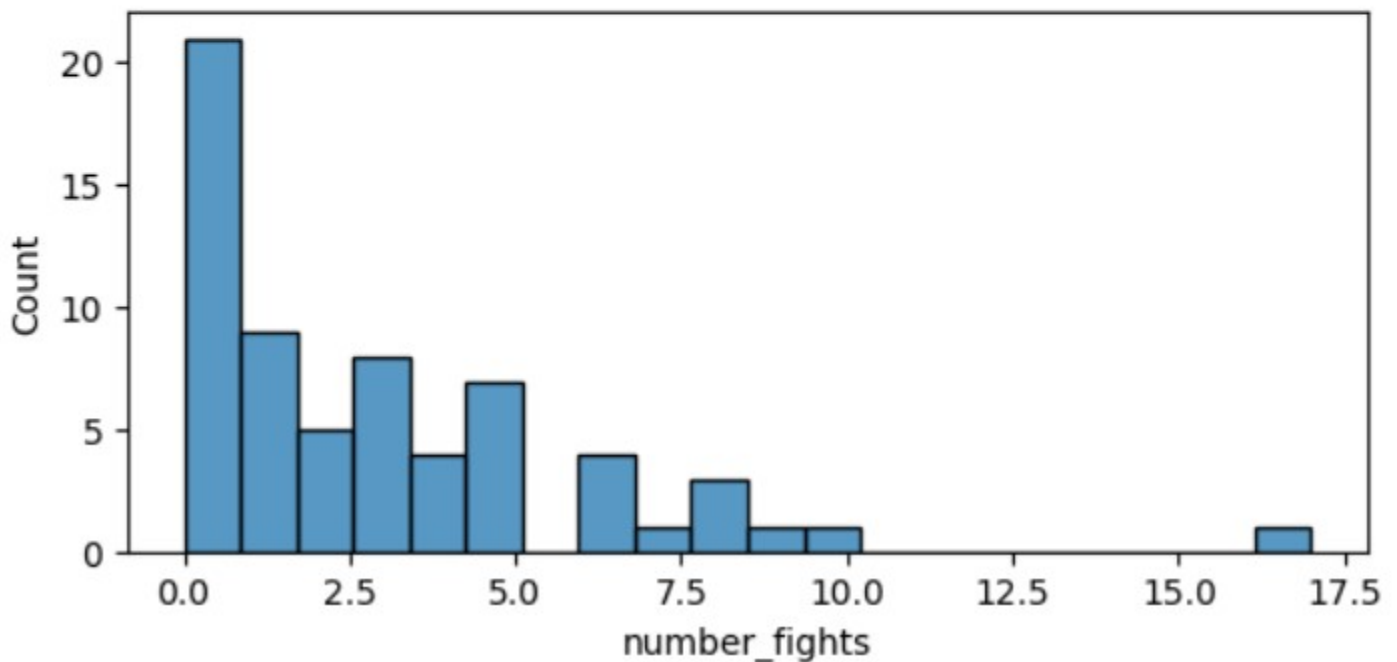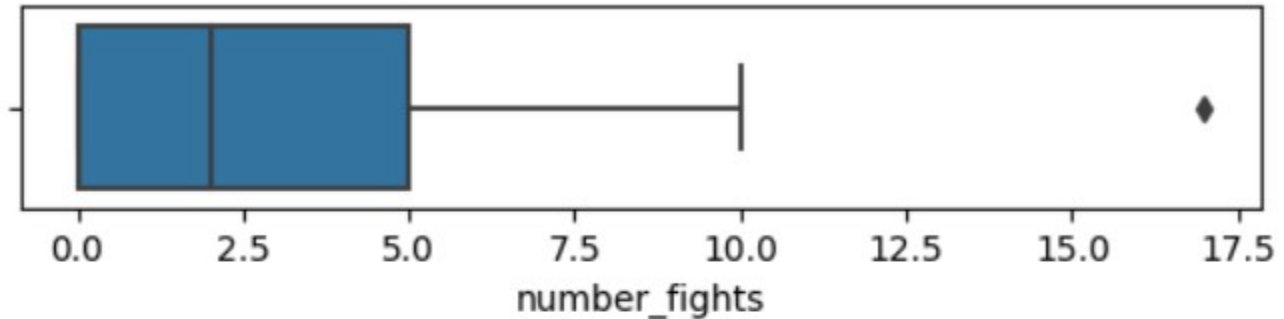# Descriptive Stats & Viz Review

1. These charts summarize a dataset about college fight songs — specifically, the number of times the word "fight" appears in each song. The box plot is missing a left whisker because $Q0 = Q1 = 0$. Using the $1.5IQR$ rule, would an observation of 16 be considered an outlier? If so, would you consider that a false positive? Explain.

2. What do these quartiles tell us about the *shape* and *spread* of the distribution? Explain.

Q0	Q1	Q2	Q3	Q4
10	21	28	37	50
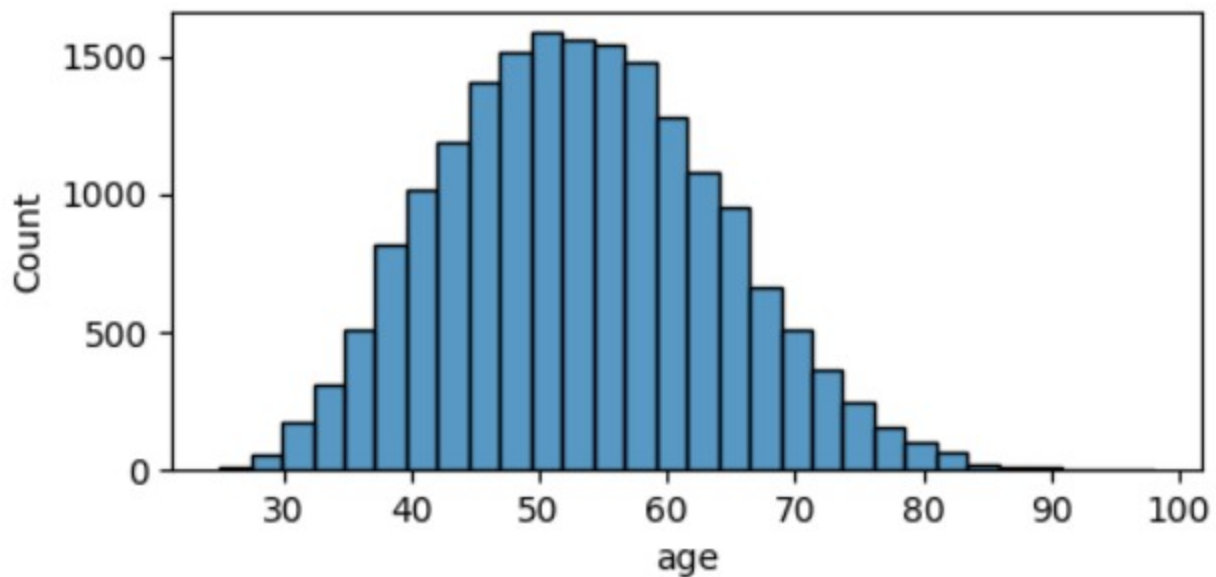
3. This histogram is built from a dataset of every member of congress who served at any point between January 1947 and February 2014. It examines the age of each congressperson. Describe this distribution. Be thorough!

4. Examine the following contingency table and fill in the conditional proportions accordingly. Based on this information, do you think there exists an association between enjoying sports and liking candy? Explain.

	Likes candy	**Dislikes candy**	**Total**
Enjoys sports	239	128	367
Does not enjoy sports	158	91	249

	Likes candy	**Dislikes candy**	**Total**	**n**
Enjoys sports			1.0	367
Does not enjoy sports			1.0	249

	state	area_name	total_employed	emp_prse	jobs_1000	location_quotient
0	PR	Aguadilla-Isabela-San Sebastian, PR	160	4.4	3.727	3.62
1	PR	Ponce, PR	180	3.6	3.434	3.34
2	PR	San German-Cabo Rojo, PR	50	0.2	2.679	2.60
3	PR	Mayaguez, PR	80	1.4	2.563	2.49
4	KY	Owensboro, KY	100	34.0	2.099	2.04
...	...	...	...	...	...	...
366	CA	Oxnard-Thousand Oaks-Ventura, CA	120	7.8	0.415	0.40
367	CA	Santa Ana-Anaheim-Irvine, CA Metropolitan Divi...	570	7.0	0.392	0.38
368	CA	Modesto, CA	60	30.1	0.359	0.35
369	CA	Bakersfield-Delano, CA	90	9.8	0.304	0.30
370	UT	Ogden-Clearfield, UT	30	15.5	0.163	0.16

371 rows × 6 columns

5. `total_employed` is the number of librarians in the respective area. Describe a visualization that would support an answer to the following question. Explain any decisions you make, including (but not limited to) marks, channels, and labels.

> How many librarians are employed in each area?

6. `location_quotient` measures the proportion of librarians to total population, specifically, per 1000 people. Describe a visualization that would support an answer to the following question. explain any decisions you make, including (but not limited to) marks, channels, and labels.

> Is there a relationship between state and location_quotient? i.e. Are librarians more common in some states than others?