MATH 262 - Applied Statistics
Lab 2: City Employee Salaries

Cesar O. Aguilar

Smallville is one of the poorest cities in Broward county and its residents have protested in the past about city officials' seemingly high salaries. You have been hired by an independent firm to investigate the employee salaries of Smallville and how they compare with the salaries of neighboring city Springfield, which is a much larger city than Smallville. The firm has provided you with a data set (`smallville_employee_salaries.csv`) that contains the 2020 salaries of all city employees of both cities. Each row in the file contains the following information about an employee:

cityName:	the name of the city in which the employee works, either Smallville or Springfield
department:	the name of the department in which the employee works
position:	the job title of the employee
wages:	the employee's salary

One of the main questions that the firm is interested in answering is whether the salaries of Smallville city managers are relatively higher than those of city managers in Springfield.

1. Create a histogram of the wages data for each city separately. Do do this in `jamovi`, use the `cityName` variable as the grouping/split by variable.

2. Is the distribution of wages for Smallville employees symmetric? If not, is it skewed to the left or to the right?

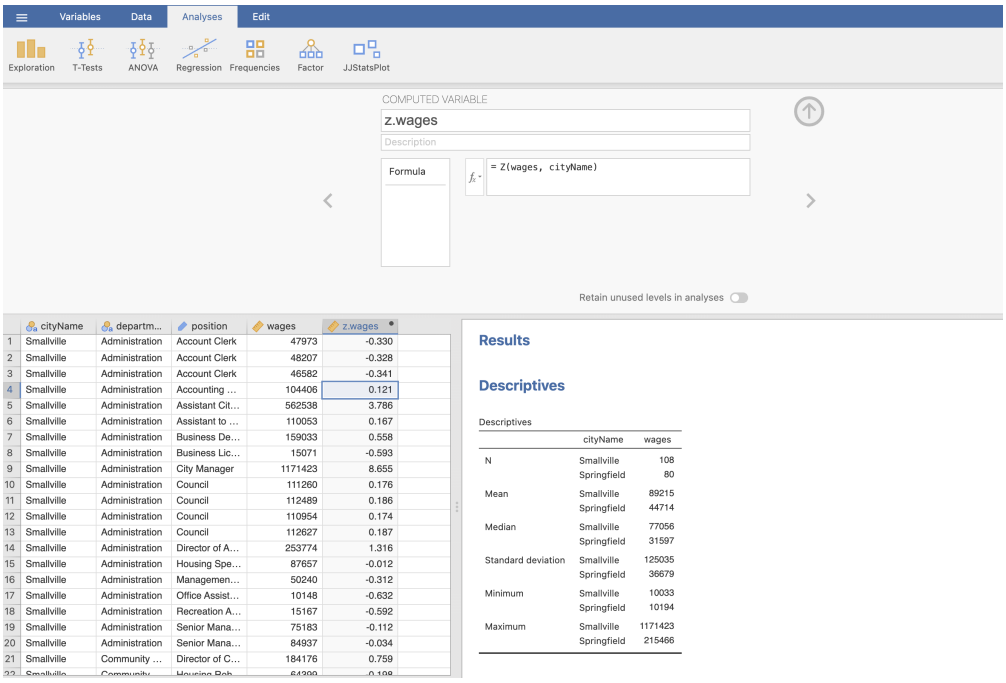
3. Without making any calculations, will the mean wage be larger or smaller than the median wage in Smallville? Explain why.

4. Now find the mean, median, and standard deviation of the wages for each city separately. Do do this in `jamovi`, use the `cityName` variable as the grouping/split by variable.

5. Which city has the larger mean wage? Based on this finding, why can you not conclude that Smallville city managers have relatively higher salaries than Springfield city managers?

6. Using a box plot, for each city determine the row number of the employee with the highest salary. What is the position of this employee for each city?

7. In most cities, the city manager is paid well above the typical city employee salary. We want to find therefore how far away the city manager salaries are from the mean for each city. In other words, we will need to compute the z-scores for the wages for each city separately. To that end, create a new **Computed Variable** in jamovi called **z.wages** using the built-in `Z` function, using the `cityName` variable as the group by variable, as shown below:



	cityName	departm...	position	wages	z.wages
1	Smallville	Administration	Account Clerk	47973	-0.330
2	Smallville	Administration	Account Clerk	48207	-0.328
3	Smallville	Administration	Account Clerk	46582	-0.341
4	Smallville	Administration	Accounting ...	104406	0.121
5	Smallville	Administration	Assistant Cit...	562538	3.786
6	Smallville	Administration	Assistant to ...	110053	0.167
7	Smallville	Administration	Business De...	159033	0.558
8	Smallville	Administration	Business Lic...	15071	-0.593
9	Smallville	Administration	City Manager	1171423	8.655
10	Smallville	Administration	Council	111260	0.176
11	Smallville	Administration	Council	112489	0.186
12	Smallville	Administration	Council	110954	0.174
13	Smallville	Administration	Council	112627	0.187
14	Smallville	Administration	Director of A...	253774	1.316
15	Smallville	Administration	Housing Spe...	87657	-0.012
16	Smallville	Administration	Managemen...	50240	-0.312
17	Smallville	Administration	Office Assist...	10148	-0.632
18	Smallville	Administration	Recreation A...	15167	-0.592
19	Smallville	Administration	Senior Mana...	75183	-0.112
20	Smallville	Administration	Senior Mana...	84937	-0.034
21	Smallville	Community ...	Director of C...	184176	0.759

	cityName	wages
N	Smallville	108
	Springfield	80
Mean	Smallville	89215
	Springfield	44714
Median	Smallville	77056
	Springfield	31597
Standard deviation	Smallville	125035
	Springfield	36679
Minimum	Smallville	10033
	Springfield	10194
Maximum	Smallville	1171423
	Springfield	215466

8. For each city, how many standard deviations is the highest salary from the mean?

9. It is common for city manager salaries to be 4 or even 5 standard deviations away from the mean. In this regard, what do you find curious about this data? Would you say that the residents of Smallville have a valid reason to protest about the city manager's salary?