

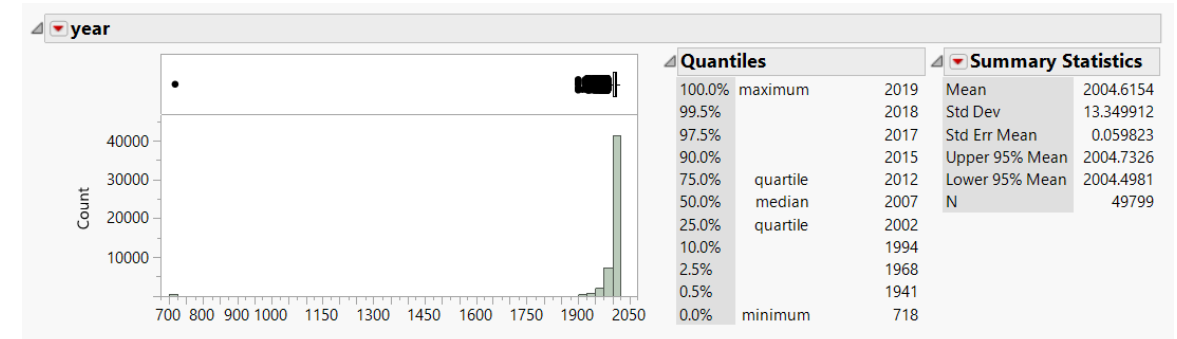
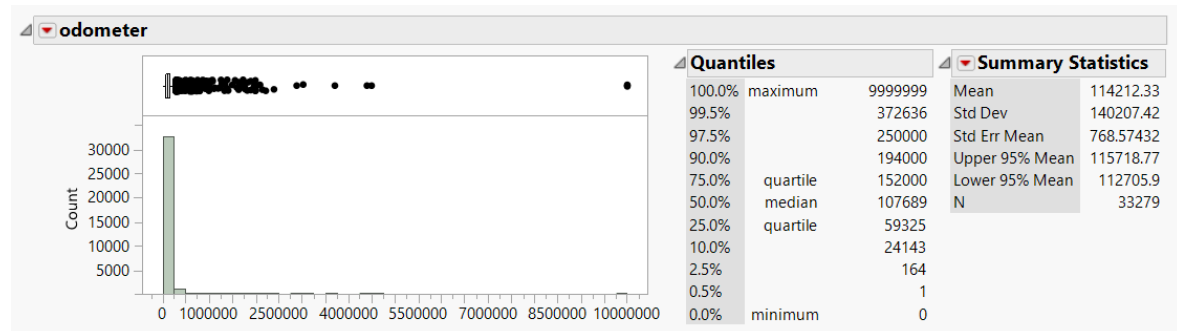
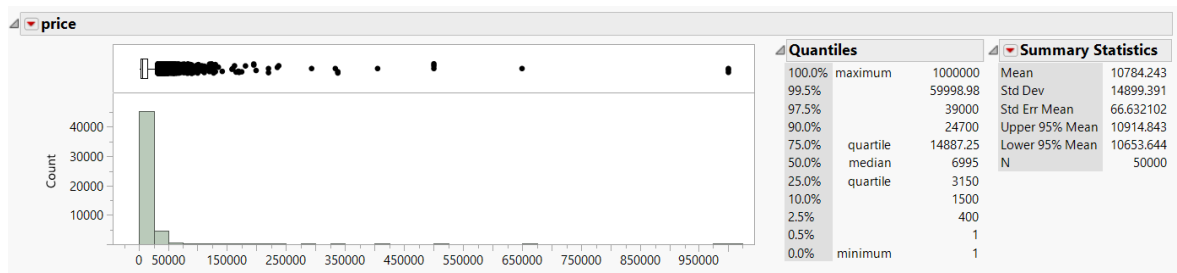
# FINAL CASE STUDY

Submitted by- Shirish Pakki

The final case relates to understanding and analysing the advertising trends related to automobile sales on Craigslist. The dataset contains a large sample of 50,000 national advertisements with various variables of the automobile like company, model, type, size, colour, transmission, title status along with geographic descriptors.

The final case analysis is meant to evaluate your skills in data cleaning, descriptive statistics and visualization and statistical data analysis using hypothesis testing for univariate and bivariate cases.

Distribution on price, odometer and year gave the following reading and all the extreme values were filtered out.



A variable 'year category' was created from Year which had 3 values- Vintage was 1970 or prior, Old was 1971 to 2000 and Modern, 2001 or later.

'year category' in table 'cleaned data 2'

Column Name:

☒ Lock

Data Type:

Modeling Type:

Column Properties ▾

Formula  
optional item

☒ Suppress Eval ☐ Ignore Errors

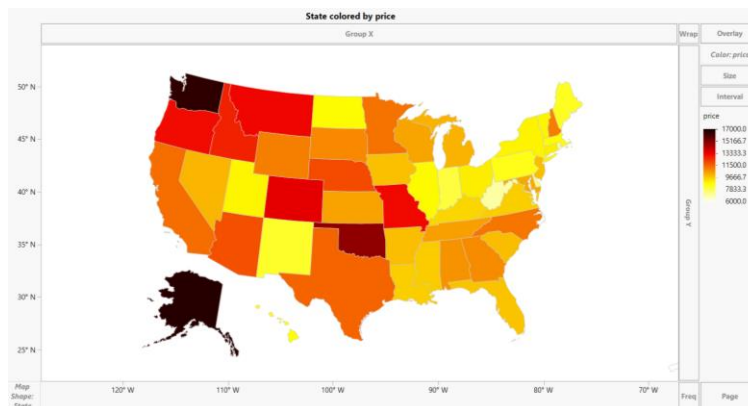
Formula

$$\text{If} \left( \begin{array}{l} \text{year} < 1971 \Rightarrow \text{"Vintage"} \\ \text{else} \Rightarrow \text{If} \left( \begin{array}{l} \text{year} < 2001 \Rightarrow \text{"Old"} \\ \text{else} \Rightarrow \text{"Modern"} \end{array} \right) \end{array} \right)$$

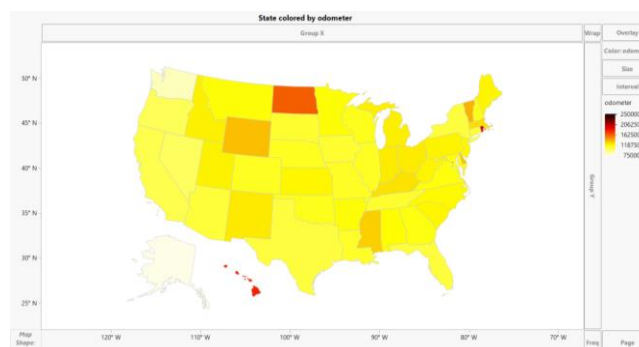
## Descriptive Analysis

a) Create visualizations depicting differences in advertising trends across the country. Use JMP's mapping facility for this step and describe your observation.

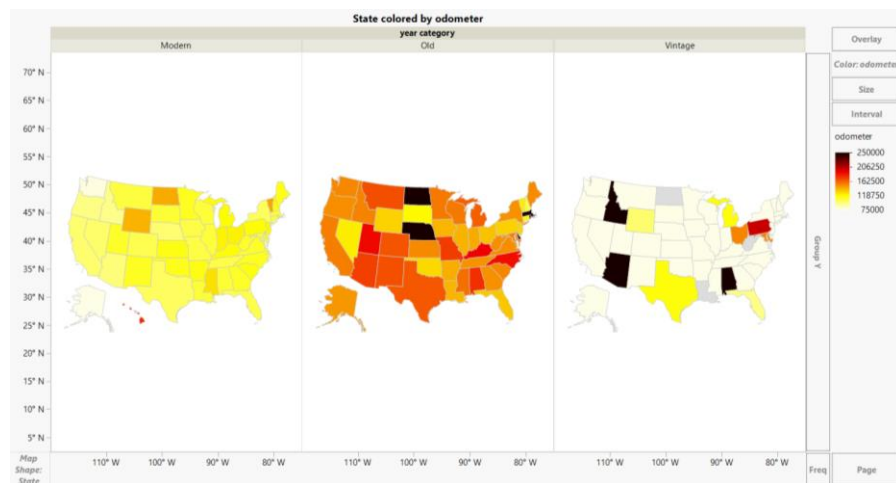
Observing price across the country, automobiles in Alaska and Washington are most expensive while West Virginia offered the cheapest vehicles. The east coast is comparatively cheaper than west coast.



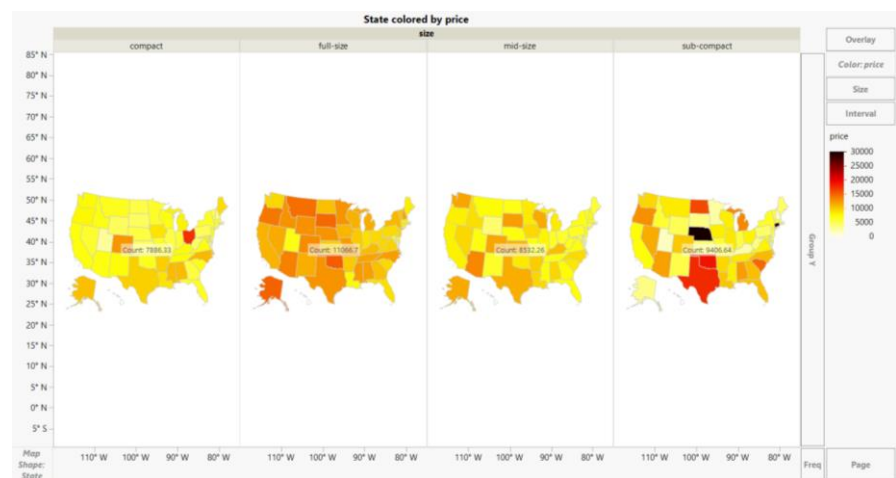
Odometer readings gave least for Alaska and Washington and highest for Hawaii. This gives an interesting observation that although the odometer readings are the lowest in Alaska and Washington, the prices are highest here too.



Vintage cars had the highest odometer readings in Idaho and Arizona while old cars with highest odometer readings were in North Dakota, Nebraska and lowest in New Hampshire. Modern cars didn't have very high odometer readings but the highest was in Hawaii and lowest Alaska and Washington.



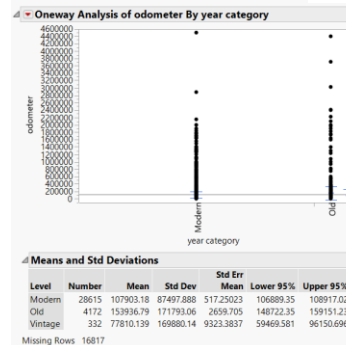
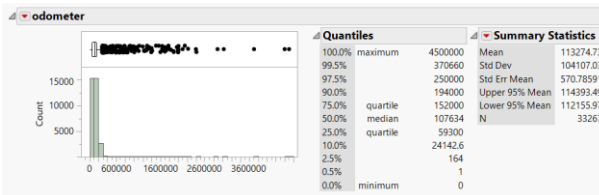
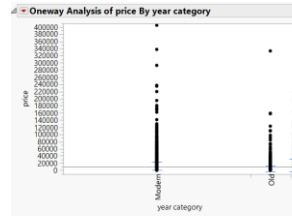
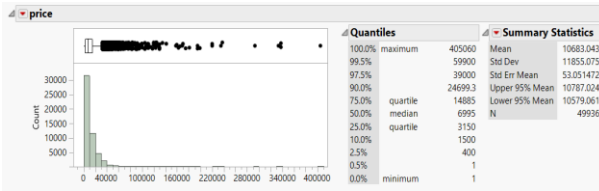
Taking price and type, Nebraska had the highest price for sub-compact vehicles while Utah had the lowest price for compact vehicles.



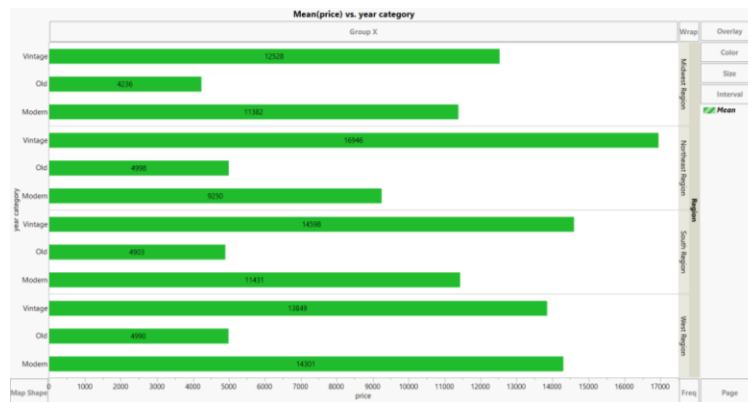
b) Run descriptive statistics for Price and Odometer variables. Now break them down by categorical Era variable (created in earlier step) and Region separately and analyse the differences. Now find a way to analyse Price and Odometer by Era and Region together to describe regional differences over time.

While the average price was \$10,683 taking each era, the mean price for vintage cars was \$14,237 while old cars priced \$4,773 and modern cars \$11,779.

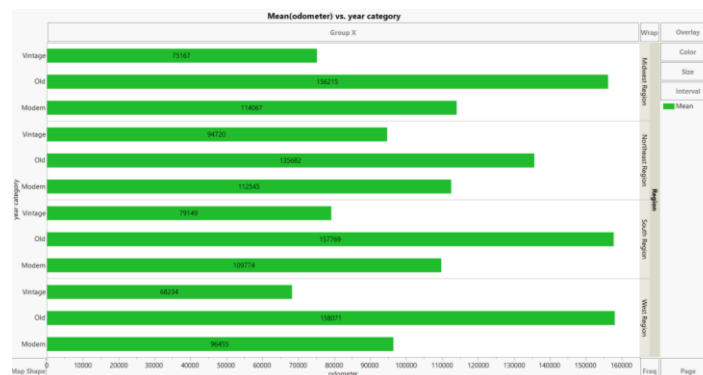
The mean odometer reading was 113,274 miles while vintage cars had 77,810 miles, old cars 153,936 and modern cars had 107,903 miles.



Taking both era and region to get mean price show vintage cars are most expensive in north east region old cars are more or less equally priced across the regions and north east had the cheapest modern cars and west had the highest.



Mean odometer reading showed lowest for vintage cars in the west and highest for northeast. The highest miles for old cars was in West and lowest in North East. Modern cars had highest miles in Midwest and lowest in West.



c) What are the most popular cars by region and division of the country?

Contingency Table		manufacturer																																																								
Count	acura	alfa	alfa-romeo	aston	audi	bmw	buick	cadillac	chevy	chevrolet	chrysler	datsun	dodge	ferarri	ford	gmc	harley	harley-davidson	honda	hyundai	infiniti	jaguar	jeep	kia	land-rover	landrover	lexus	lincoln																														
Total %	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00																														
Col %	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00																														
Row %	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00																														
Midwest Region	93	0	1	0	85	198	280	240	5	2055	392	310	3	649	1	4	2440	640	1	5	509	186	59	1	18	981	161	0	0	87																												
	0.20	0.00	0.00	0.00	0.18	0.43	0.61	0.52	0.01	4.46	0.85	0.67	0.01	1.41	0.00	0.01	5.30	1.41	0.00	0.01	1.11	0.40	0.13	0.00	0.04	1.26	0.35	0.00	0.00	0.19																												
	1670	0.00	100.00	0.00	1895	15.68	42.62	22.35	18.52	31.89	33.14	34.22	15.00	29.57	16.67	9.30	28.41	30.83	11.11	23.33	19.74	20.20	17.99	4.00	14.63	26.07	23.85	0.00	0.00	14.01																												
	0.79	0.00	0.01	0.00	0.73	1.69	2.40	2.05	0.94	11.78	3.36	2.65	0.03	1.56	0.01	0.03	10.88	1.56	0.01	0.04	4.36	1.59	0.51	0.01	0.15	4.97	1.38	0.00	0.00	9.74																												
Northeast Region	123	1	0	1	96	212	88	99	3	783	172	123	3	302	0	5	1093	247	3	0	486	185	47	5	17	382	82	1	1	87																												
	0.27	0.00	0.00	0.00	0.21	0.46	0.19	0.22	0.01	3.70	0.37	0.27	0.01	0.66	0.00	0.01	2.37	0.54	0.01	0.00	1.06	0.40	0.10	0.01	0.04	0.83	0.18	0.00	0.00	0.19																												
	22.32	50.00	0.00	33.33	22.54	16.79	13.39	13.54	11.11	12.15	14.54	13.58	15.00	13.76	0.00	11.63	12.75	11.73	33.33	0.00	18.85	20.09	14.33	20.00	13.82	17.14	12.15	10.00	50.00	14.01																												
	1.82	0.01	0.00	0.01	1.42	3.14	1.30	1.46	0.04	11.58	2.34	1.82	0.04	4.47	0.00	0.07	16.11	3.65	0.04	0.00	7.19	2.74	0.70	0.07	0.25	5.65	1.21	0.01	0.01	1.29																												
South Region	168	0	0	1	106	400	191	239	6	2326	384	287	7	719	1	3	1952	465	2	5	951	361	141	12	17	174	284	0	1	248																												
	0.41	0.00	0.00	0.00	0.23	0.86	0.41	0.52	0.01	5.05	0.86	0.62	0.02	1.56	0.00	0.05	6.41	1.48	0.00	0.01	2.07	0.78	0.31	0.03	0.12	1.59	0.62	0.00	0.00	0.53																												
	34.12	0.00	0.00	33.33	24.88	35.63	29.07	22.21	22.22	36.09	33.31	31.68	35.00	32.76	16.67	48.84	34.44	32.40	22.22	23.33	36.89	39.20	42.99	48.00	46.34	32.93	42.07	0.00	50.00	39.61																												
	1.16	0.00	0.00	0.01	0.66	2.78	1.18	1.48	0.04	14.39	2.44	1.78	0.04	4.45	0.01	0.11	18.97	4.22	0.01	0.03	5.88	2.23	0.87	0.07	0.35	4.54	1.76	0.00	0.01	1.52																												
West Region	148	1	0	1	139	493	98	164	13	1281	225	186	7	525	4	13	2086	527	3	5	632	189	81	7	31	532	148	0	6	201																												
	0.32	0.00	0.00	0.00	0.30	1.08	0.21	0.36	0.03	3.78	0.49	0.40	0.03	1.14	0.01	0.03	4.53	1.14	0.01	0.01	1.37	0.41	0.18	0.02	0.07	1.16	0.32	0.00	0.00	0.44																												
	26.86	50.00	0.00	33.33	32.63	31.91	14.92	22.10	48.15	19.88	19.02	20.33	35.00	23.92	66.67	30.23	24.54	25.04	33.33	33.33	24.52	20.52	24.70	28.00	25.20	23.87	21.93	0.00	0.00	32.37																												
	1.29	0.01	0.00	0.01	1.21	3.52	0.86	1.43	0.11	11.20	1.97	1.63	0.06	4.39	0.03	0.11	16.23	4.61	0.03	0.04	5.52	1.65	0.71	0.06	0.27	4.65	1.29	0.00	0.00	1.76																												
	551	2	1	3	426	1363	657	742	27	6445	1183	906	20	2195	6	43	4971	2105	9	15	2719	921	328	25	121	2329	675	1	2	421																												
Total	1.20	0.00	0.00	0.01	0.93	2.74	1.43	1.61	0.06	14.00	2.57	1.97	0.04	4.71	0.01	0.09	18.61	4.57	0.02	0.03	5.60	2.00	0.71	0.05	0.27	4.84	1.47	0.00	0.00	1.35																												
Tests																																																										
N	46046																																																									
DF	144																																																									
-Loglike	10295.5094																																																									
R Square (R2)	0.0075																																																									
Test																																																										
ChiSquare																																																										
Prob>ChiSq																																																										
Likelihood Ratio		2059.001																																																								
Pearson		2048.962																																																								
		<0.0001*																																																								
		Warning: 20% of cells have expected count less than 5. ChiSquare suspect.																																																								

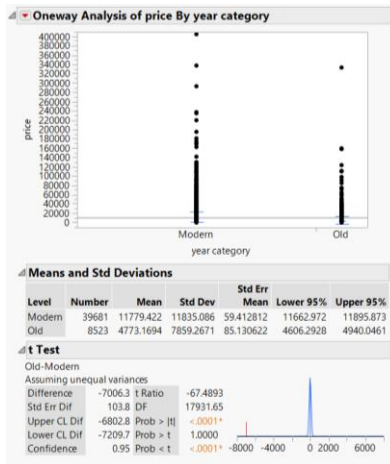
The highest number of cars on Craigslist appear to be Fords irrespective of region and division followed by Chevrolet.

Contingency Table		manufacturer																												
Count	acura	alfa	alfa-romeo	aston	audi	bmw	buick	cadillac	chevrolet	chevrolet	chrysler	datsun	dodge	ferarri	ford	ford	gmc	harley	harley-davidson	honda	hyundai	infiniti	jaguar	jeep	kia	land-rover	landrover	lexus		
Total %	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
Col %	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
Row %	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
East North	58	0	0	0	50	131	188	159	3	1212	237	200	1	416	1	3	1475	405	1	3	327	132	40	1	16	384	111	0	0	
	0.13	0.00	0.00	0.00	0.11	0.28	0.43	0.35	0.01	2.63	0.51	0.43	0.00	0.90	0.00	0.01	3.20	0.88	0.00	0.01	0.71	0.29	0.09	0.00	0.03	0.86	0.24	0.00	0.00	
	10.71	0.00	0.00	0.00	11.74	10.37	30.14	21.43	11.11	18.81	20.03	23.08	5.00	18.94	16.67	6.98	17.23	19.24	11.11	20.00	13.68	14.33	12.20	4.00	13.01	17.68	16.44	0.00	0.00	
	0.81	0.00	0.00	0.00	0.68	1.79	2.71	2.18	0.04	10.19	3.24	2.74	0.01	3.70	0.01	0.04	10.19	5.54	0.01	0.04	4.48	1.81	0.55	0.01	0.22	5.39	1.52	0.00	0.00	
East South	23	0	0	0	18	51	47	37	1	444	87	37	3	126	0	3	323	148	0	0	129	50	22	1	7	124	54	0	1	
	0.05	0.00	0.00	0.00	0.04	0.11	0.10	0.08	0.00	0.94	0.19	0.08	0.01	0.27	0.00	0.01	1.15	0.32	0.00	0.00	0.28	0.11	0.05	0.00	0.02	0.27	0.12	0.00	0.00	
	4.34	0.00	0.00	0.00	4.23	4.04	7.15	4.99	3.70	6.15	7.35	4.08	15.00	5.74	0.00	6.98	6.16	7.03	0.00	0.00	5.65	6.71	4.00	0.04	5.69	5.56	8.00	5.00	7.00	
	0.88	0.00	0.00	0.00	0.63	1.86	1.66	1.26	0.04	2.07	1.20	0.11	4.44	0.11	0.07	3.88	5.22	0.00	0.00	4.83	1.83	0.78	0.04	0.22	4.37	1.90	0.04	0.00	1.00	
Central Division	76	1	0	0	54	145	64	66	2	544	126	93	3	208	0	1	766	147	0	0	299	126	29	4	2	251	60	1	0	
	0.17	0.00	0.00	0.00	0.12	0.31	0.14	0.15	0.00	1.08	0.27	0.20	0.01	0.45	0.00	0.01	1.62	0.32	0.00	0.00	0.65	0.27	0.07	0.01	0.02	0.55	0.13	0.00	0.00	
	13.79	0.00	0.00	0.00	13.79	34.00	14.44	15.54	0.04	10.44	2.66	2.08	0.07	4.65	0.00	0.07	17.12	3.38	0.00	0.00	6.88	2.82	0.65	0.09	0.14	5.41	1.34	0.02	0.01	
	1.70	0.02	0.00	0.02	1.21	3.24	1.43	1.54	0.04	12.36	3.22	2.08	0.07	4.65	0.00	0.07	17.12	3.38	0.00	0.00	6.88	2.82	0.65	0.09	0.14	5.41	1.34	0.02	0.01	
Mountain Division	7	1	0	0	4	11	10	13	0	119	23	10	0	19	0	0	64	12	0	0	21	9	3	0	0	1	1	0	0	0
	0.00	0.00	0.00	0.02	0.12	0.19	0.09	0.04	0.01	1.14	0.20	0.16	0.00	0.46	0.00	0.01	1.60	0.49	0.00	0.00	0.40	0.16	0.05	0.00	0.02	0.06	0.13	0.00	0.00	
	0.08	0.00	0.00	0.02	0.14	0.27	0.14	0.18	0.01	1.71	0.36	0.16	0.00	0.64	0.00	0.01	2.00	0.64	0.00	0.00	0.50	0.21	0.04	0.00	0.01	0.13	0.16	0.00	0.00	
	6.72	0.00	0.00	0.24	6.97	6.24	8.68	11.11	0.11	7.61	0.86	0.50	0.00	9.74	0.30	0.81	10.89	0.00	0.00	7.18	7.93	7.32	8.00	6.50	1.48	0.19	0.93	0.00	0.00	
	0.08	0.00	0.00	0.03	0.15	0.07	0.08	0.07	0.00	0.12	0.10	0.06	0.00	0.12	0.00	0.01	0.13	0.32	0.00	0.00	0.88	0.91	0.84	0.00	0.23	0.12	0.00	0.00	0.00	
New England Division	47	0	0	0	42	67	24	30	1	239	46	30	0	94	0	0	328	100	0	0	187	59	18	1	10	131	22	0	0	
	0.11	0.00	0.00	0.00	0.09	0.15	0.06	0.07	0.00	0.52	0.07	0.07	0.00	0.21	0.00	0.00	0.74	0.23	0.00	0.00	0.43	0.13	0.04	0.00	0.03	0.28	0.06	0.00	0.00	
	8.50	0.00	0.00	0.00	8.46	12.06	4.46	5.30	0.17	3.89	0.60	0.40	0.00	1.68	0.00	0.00	5.75	2.05	0.00	0.00	3.25	1.04	0.31	0.01	1.31	1.26	0.00	0.00	0.00	
	2.05	0.00	0.00	0.00	1.84	2.59	1.05	1.31	0.04	15.45	2.01	1.31	0.00	4.11	0.00	0.00	14.84	4.37	0.00	0.00	8.17	2.58	0.79	0.04	0.44	5.73	0.96	0.00	0.00	
	3.56	0.00	0.00	0.00	3.31	4.61	1.91	2.51	0.13	10.13	1.31	0.87	0.00	3.08	0.00	0.00	14.84	4.37	0.00	0.00	8.17	2.58	0.79	0.04	0.44	5.73	0.96	0.00	0.00	
Pacific Division	4	0	0	0	3	10	13	13	0	101	13	10	0	19	0	0	64	12	0	0	21	9	3	0	0	1	1	0	0	0
	0.24	0.00	0.00	0.00	0.19	0.68	0.12	0.02	0.02	1.65	0.29	0.25	0.01	0.68	0.01	0.02	2.39	0.68	0.01	0.01	0.97	0.25	0.12	0.01	0.04	0.50	0.09	0.19	0.00	
	20.14	0.00	0.00	0.33	21.19	24.94	8.68	1.61	37.04	11.14	11.27	3.00	0.00	12.41	0.68	0.01	27.53	14.35	0.33	13.33	17.40	17.38	20.00	18.70	1.28	12.74	0.00	0.00	0.00	
	15.31	0.00	0.00	0.01	13.31	15.00	5.00	0.04	16.67	5.00	5.00	1.00	0.00	6.67	0.00	0.00	16.67	8.33	0.33	13.33	17.40	17.38	20.00	18.70	1.28	12.74	0.00	0.00	0.00	
South Atlantic Division	2	0	0	0	6	28	92	1	3	1189	166	100	1	378	0	0	9	1352	317	2	3	962	286	83	9	37	431	147	0	1
	0.27	0.00	0.00	0.00	0.20	0.20	0.20	0.01	0.24	0.36	0.39	0.02	0.01	0.98	0.00	0.00	0.01	1.39	0.40	0.16	0.00	0.36	0.36	0.04	0.56	0.32	0.00	0.00	0.00	
	22.50	0.00	0.00	0.00	10.00	22.64	14.07	0.13	11.71	18.14	14.03	19.87	5.00	17.21	0.67	26.93	18.11	15.96	22.22	20.20	22.96	25.44	23.30	36.00	19.34	21.78	0.70	0.20	0.00	
	1.46	0.00	0.00	0.00	0.68	3.24	1.04	1.32	0.02	13.22	1.88	2.00	0.01	4.28	0.01	0.00	1.75	3.39	0.00	0.00	2.76	2.56	3.94	0.10	0.42	4.88	1.86	0.00	0.00	
	0.07	0.00	0.00	0.00	0.12	0.47	0.17	0.11	0.00	0.24	0.04	0.04	0.00	0.11	0.00	0.00	0.04	0.08	0.00	0.00	0.13	0.12	0.04	0.00	0.01	0.11	0.00	0.00	0.00	
West North	599	0	0	0	822	1336	1052	741	0	6546	1330	24	0	1021	0	0	2321	1139	0	0	1586	579	140	0	0	140	741	0	0	0
	0.07	0.00	0.00	0.00	0.08	0.15	0.18	0.10	0.00	1.83	0.34	0.24	0.00	0.51	0.00	0.00	2.20	0.53	0.00	0.00	0.40	0.12	0.04	0.00	0.00	0.41	0.11	0.00	0.00	
	10.78	0.00	0.00	0.00	8.22	12.46	10.52	7.41	0.00	115.46	23.30	24	0.00	10.61	0.00	0.00	40.31	23.21	0.00	0.00	27.56	5.79	1.40	0.00	0.00	2.41	1.27	0.00	0.00	0.00
	0.75	0.00	0.00	0.00	0.80	1.53	1.87	1.85	0.00	19.25	3.54	2.51	0.00	5.32	0.00	0.00	22.04	11.39	0.00	0.00	4.16	1.23	0.41	0.00	0.00	0.42	1.14	0.00	0.00	0.00
West South	39	0	0	0	28	113	52	85	2	723	141	70	0	215	0	0	9	872	217	2	2	280	83	36	2	13	179	83	0	0
	0.09	0.00	0.00	0.00	0.07	0.28	0.12	0.18	0.00	1.40	0.31	0.16	0.00	0.48	0.00	0.00	0.02	0.24	0.00	0.00	0.40	0.12	0.09	0.05	0.03	0.31	0.21	0.00	0.00	0.00
	7.08	0.00	0.00	0.00	3.33	6.57	8.95	7.91	11.46	7.41	11.22	11.92	7.33	15.00	9.79	0.00	20.93	10.17	0.01	0.01	13.33	8.92	10.96	8.00	10.57	8.03	12.30	0.00	0.00	
	0.08	0.00	0.00	0.00	0.07	0.28	0.12	0.18	0.00	1.40	0.31	0.16	0.00	0.48	0.00	0.00	0.02	0.24	0.00	0.00	0.40	0.12	0.09	0.05	0.03	0.31	0.21	0.00	0.00	0.00
Total	611	1	0	0	422	1362	675	317	5	6493	1330	24	0	1021	0	0	2321	1139	0	0	1586	579	140	0	0	140	741	0	0	0

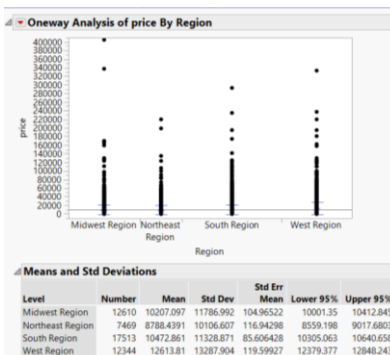
## Hypothesis Testing

Run the following tests of hypothesis and provide relevant output and analysis in your own words.

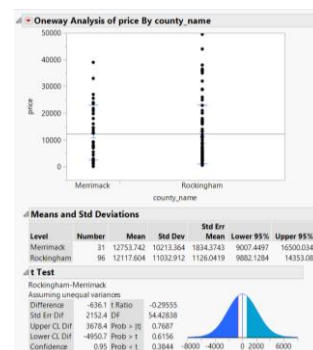
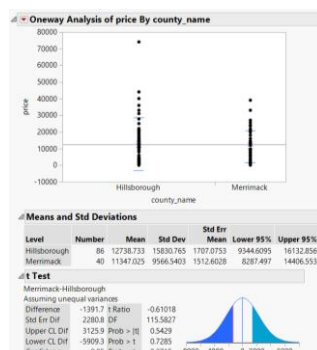
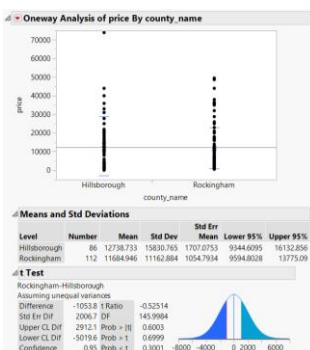
- a) Are prices different by era? Is there a statistically significant difference in list price of Old and Modern automobiles? Does the difference hold at the same level when broken down by region (compare confidence intervals of the price difference)?



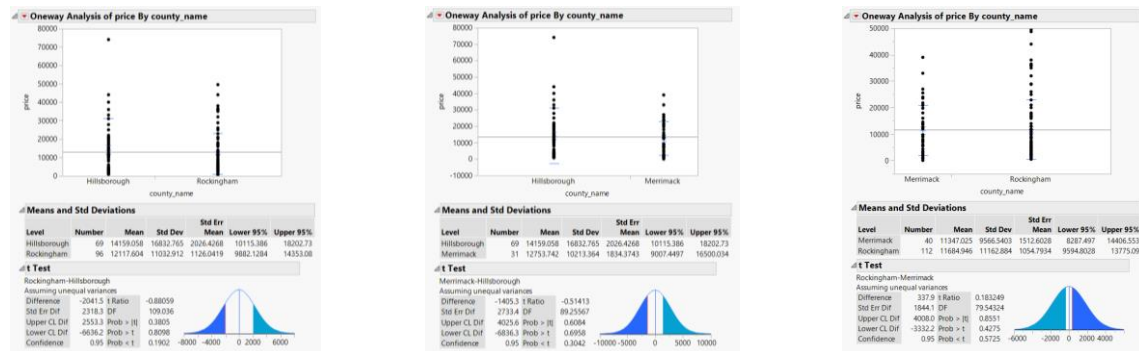
Old cars were cheaper than modern cars. Comparing the regions, Midwest and south cannot be compared as their confidence intervals overlap. Northeast regions offered the cheapest cars and west had highest.



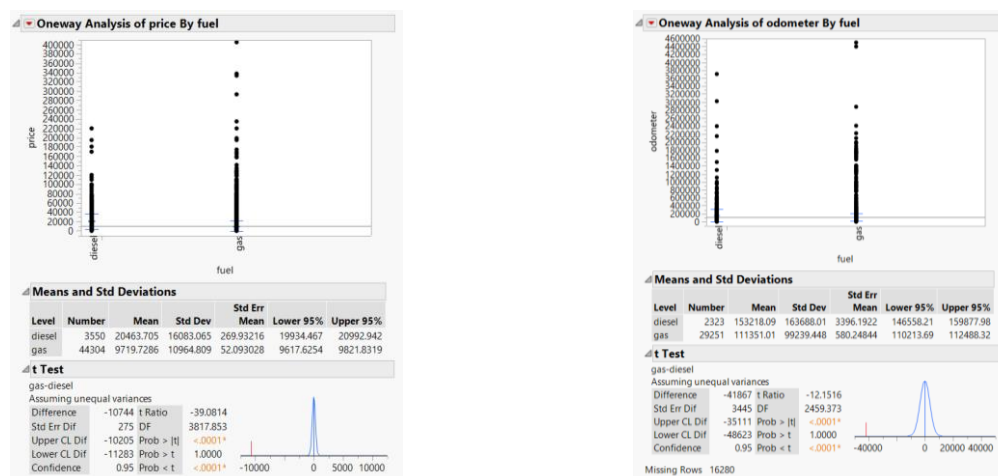
- b) Within state differences: Hillsborough, Rockingham and Merrimack are the three biggest counties in NH. You are asked to analyse if there are differences in the kinds of cars being advertised in these counties. Run three pairwise tests of comparison of mean price for all eras for these counties. Now run the same tests just for Modern cars. Are the conclusions in the two kinds of tests different?



The means are too close to each other and so, it is not possible to compare the values. Also, the p-values are all high so all theories have failed to be rejected.



c) Difference in price by Fuel type: Compare Price and Odometer reading of Gas and Diesel vehicles using hypothesis testing and describe your findings. Now run the same tests for Northeast and South regions separately. Are the findings still the same? If not, then describe the differences in your own words.



Diesel engines are expensive than gas engines. Diesel engines also appear to have higher miles than gas engines. The same is the case for northeast and south regions.

