Car Sales EDA FInal Edition

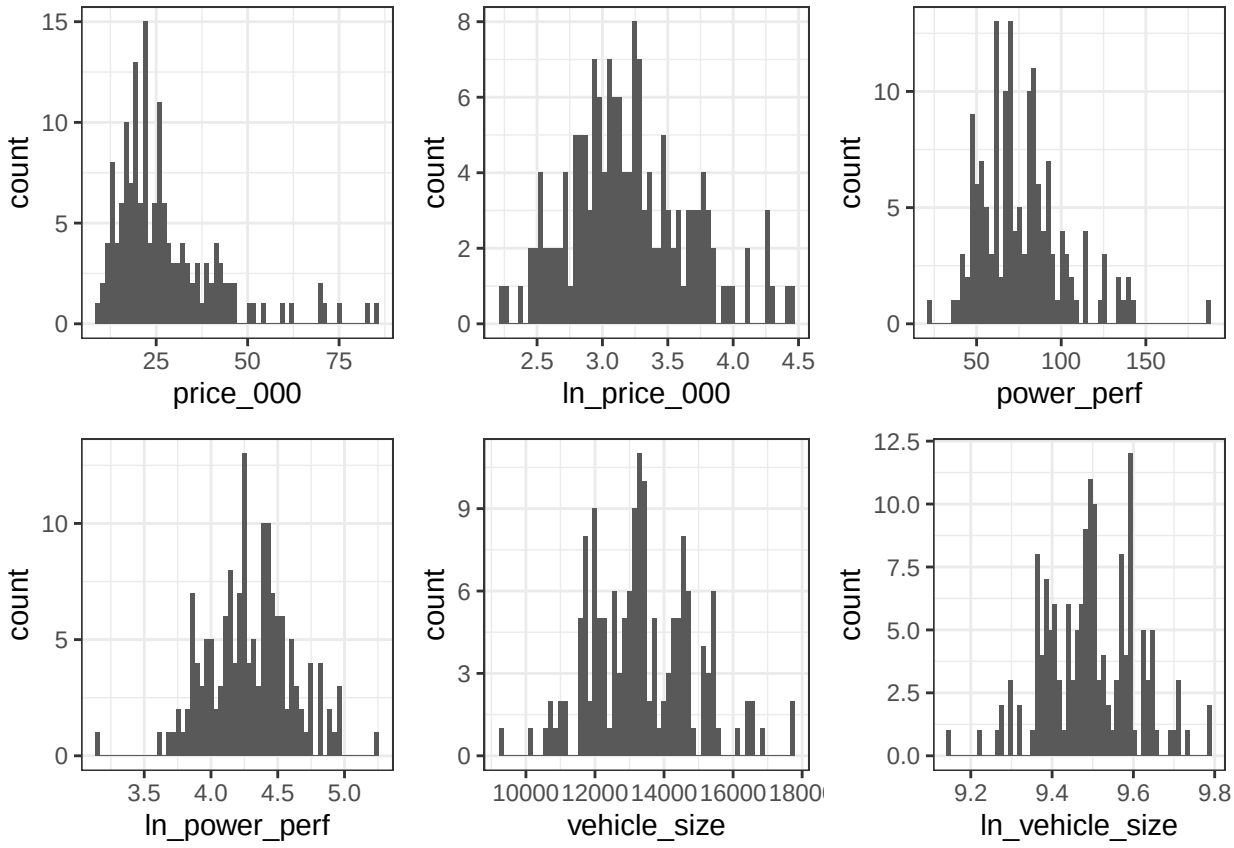
Ed Brown, Daphne Lin, Linh Tran, Lisa Wu

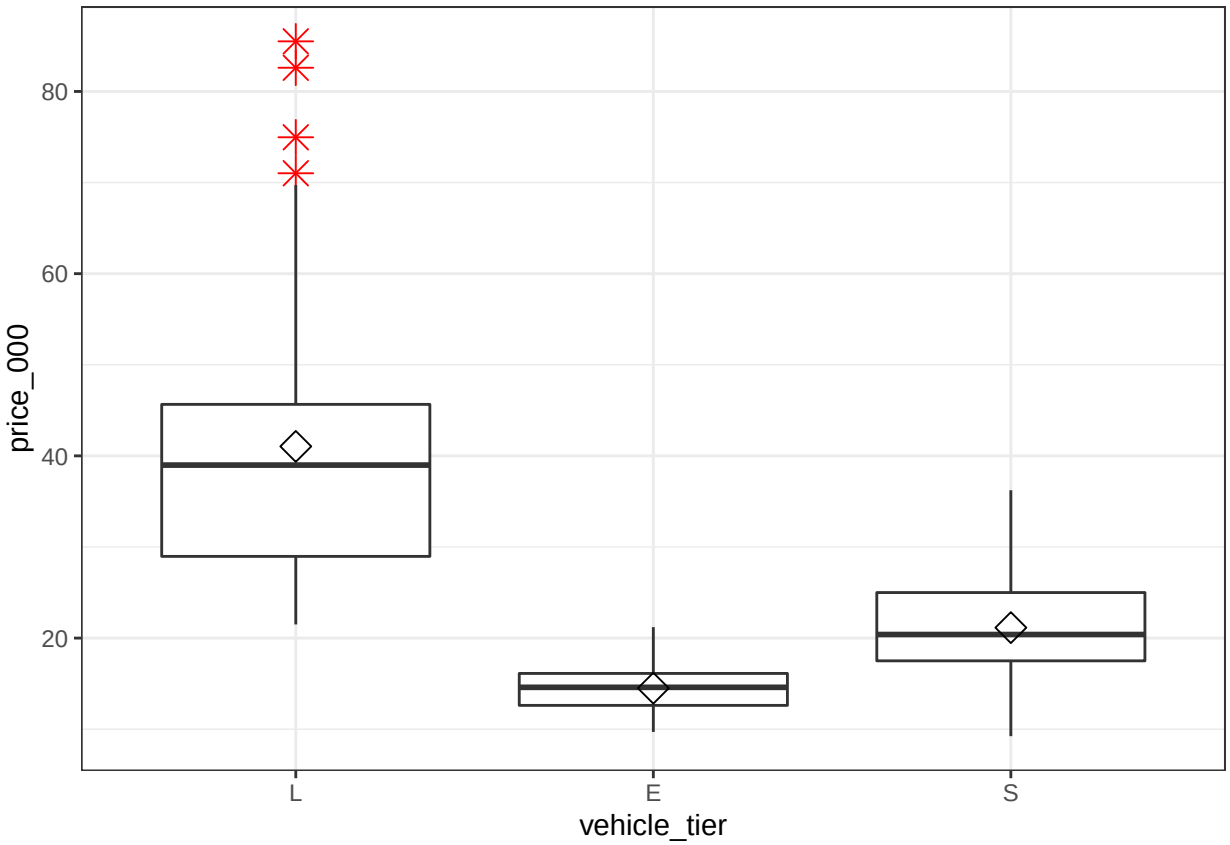
2022-07-23

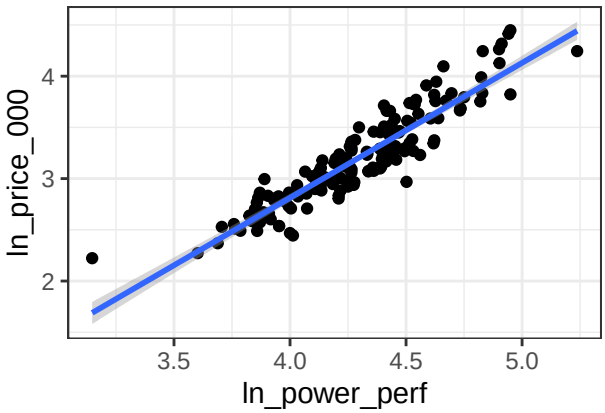
```
# Remove one row "Town & Country" has no values
car_sales <- car_sales[!(car_sales$model == "Town & Country"), ]

# Remove all rows with missing power_performance information (1 row)
car_sales <- car_sales[!is.na(car_sales$power_perf), ]
```

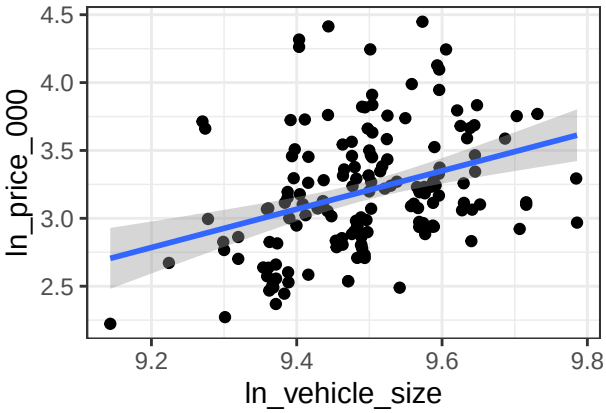
```
# Convert to Factors
car_sales$manufacturer <- as.factor(car_sales$manufacturer)
car_sales$model <- as.factor(car_sales$model)
car_sales$vehicle_type <- as.factor(car_sales$vehicle_type)
# Vehicle Tier L = Luxury; S = Standard; E = Economy
# Convert to Factor and Re-Level to Luxury as Base Model
car_sales[, "vehicle_tier"] <- relevel(as.factor(car_sales$vehicle_tier),
                                       ref = "L")
# Convert Dates
car_sales$latest_launch <- as.POSIXct(car_sales$latest_launch,
                                      format = "%m/%d/%Y")
```

```
car_sales <- car_sales %>%
  mutate(
    ln_price_000 = log(price_000),
    ln_width = log(width),
    ln_curb_weight = log(curb_weight),
    ln_fuel_capacity = log(fuel_capacity),
    ln_fuel_efficiency = log(fuel_efficiency),
    ln_power_perf = log(power_perf),
    vehicle_range = fuel_capacity * fuel_efficiency,
    ln_range = log(vehicle_range),
    vehicle_size = length * width,
    ln_vehicle_size = log(vehicle_size),
    density = curb_weight / ln_vehicle_size,
    ln_density = log(density),
    days_since_refresh = as.numeric(difftime(as.POSIXct(Sys.Date(), tz = "UTC"),
                                             latest_launch, units = "days"))
  )
car_sales$refresh_normalized <-
  car_sales$days_since_refresh / max(car_sales$days_since_refresh)
```


count
price_000
ln_price_000
power_perf
ln_power_perf
vehicle_size
ln_vehicle_size


80
60
40
20
price_000
L
E
S
vehicle_tier


ln_price_000
4
3
2
3.5
4.0
4.5
5.0
ln_power_perf


ln_price_000
4.5
4.0
3.5
3.0
2.5
9.2
9.4
9.6
9.8
ln_vehicle_size

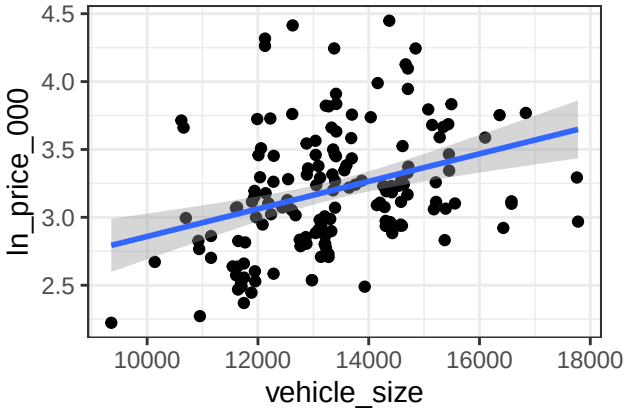

ln_price_000
4.5
4.0
3.5
3.0
2.5
10000
12000
14000
16000
18000
vehicle_size

Table 1: Estimated Regressions

	Output Variable: natural log of price in thousands of dollars		
	(1)	(2)	(3)
Natural Log Power Performance Ratio	1.318^{***}	1.470^{***}	1.210^{***}
	(0.061)	(0.061)	(0.074)
Economy Tier Vehicles			-0.256^{***}
			(0.056)
Mid Tier Vehicles			-0.213^{***}
			(0.032)
Vehicle Size		-0.812^{***}	-0.550^{***}
		(0.153)	(0.148)
constant	-2.458^{***}	4.595^{***}	3.369^{**}
	(0.259)	(1.356)	(1.236)
Observations	155	155	155
R^2	0.846	0.875	0.904
Adjusted R^2	0.845	0.873	0.901
Residual Std. Error	0.179 (df = 153)	0.162 (df = 152)	0.143 (df = 150)
F Statistic	841.934^{***} (df = 1; 153)	530.570^{***} (df = 2; 152)	352.724^{***} (df = 4; 150)
Note:	*p<0.05; **p<0.01; ***p<0.001		
	HC_1 robust standard errors in parentheses. Luxury Vehicles are the base Tier		