Deals on Wheels: let the market show you how to buy a better car

Michael Boles

February 2020

Executive summary

Introduction

Methods

**Results**

1. *Exploratory data analysis*

Approximately 90,000 currently listed new and used cars were scraped from the Autotrader web page in January 2020 across five major metro areas (New York, Los Angeles, Chicago, Houston, and San Francisco). Features collected include price, make, model, year, mileage, location, body style, engine, transmission, and drive type.

The distribution of list prices across the set (Figure 1, left) displays a long tail, with a mode of $7,000 and median price of $16,000. Reported mileage is trimodal: a large number of new cars with zero miles (20,000 listings), a second set with approximately 30,000 miles, and a third set with about 100,000 miles. Given that the typical car is driven something like 10,000 miles per year (Figure 1, right), that middle segment may reflect listings of leased cars (typically with 24- to 48-month lease terms) appearing on the market.

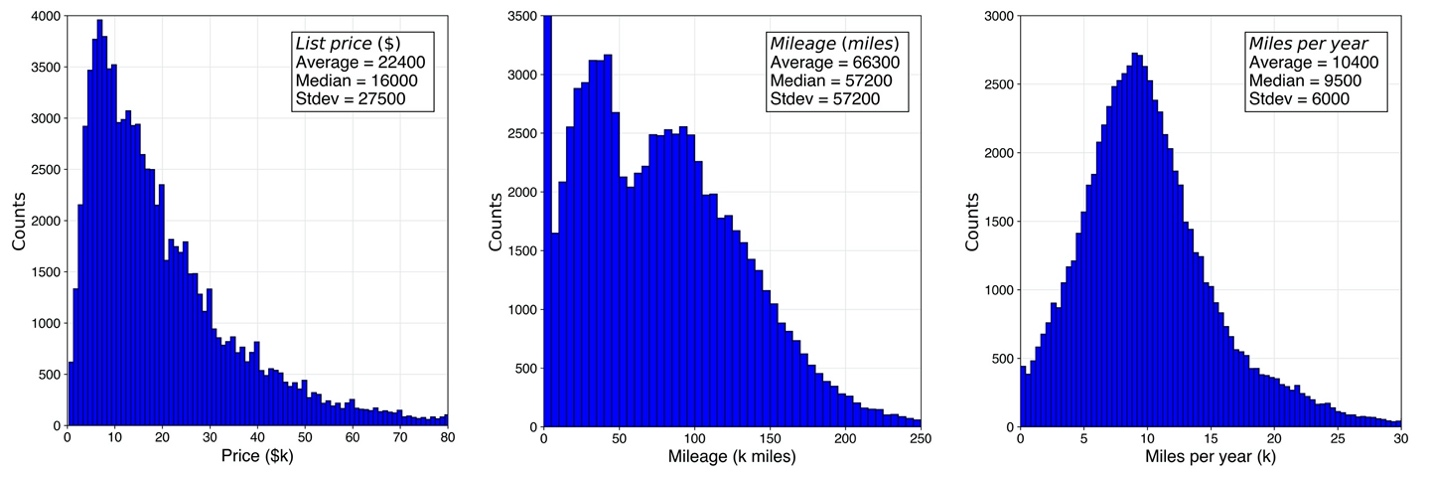


Figure 1. Histograms of (left to right) price, mileage, and miles per year across the 100,000 listings collected from Autotrader in January 2020.

Sales of sport utility vehicles (SUVs) were reported to surpass those of sedans in 2014, (<https://www.edmunds.com/car-news/sedan-dethroned-as-most-popular-body-style-in-america.html>) and this is indeed reflected in the Autotrader listings (Figure 2, left). Together, SUVs and sedans make up more than half of cars on the market. Minivans, which were three times more popular in the year 2000 than they are now (<https://www.freep.com/story/money/cars/2019/08/02/minivan-sales/1898974001/>), are beside station wagons at the bottom of the list.

Despite retaining just 10% of the US market, the truck segment lays claim to the most popular car in America, the Ford F150 (Figure 2, right). The top 10 most frequently encountered cars in this data set include 4 sedans (Toyota and Honda’s midsize- and small offerings of Camry, Accord, Civic, and Corolla), 4 SUVs (Jeep’s Grand Cherokee and Wrangler, together with the Honda CR-V and Toyota Highlander), and 2 trucks (the aforementioned Ford F150 and Chevrolet Silverado 1500). One luxury offering made it into the top 20: the BMW X5.

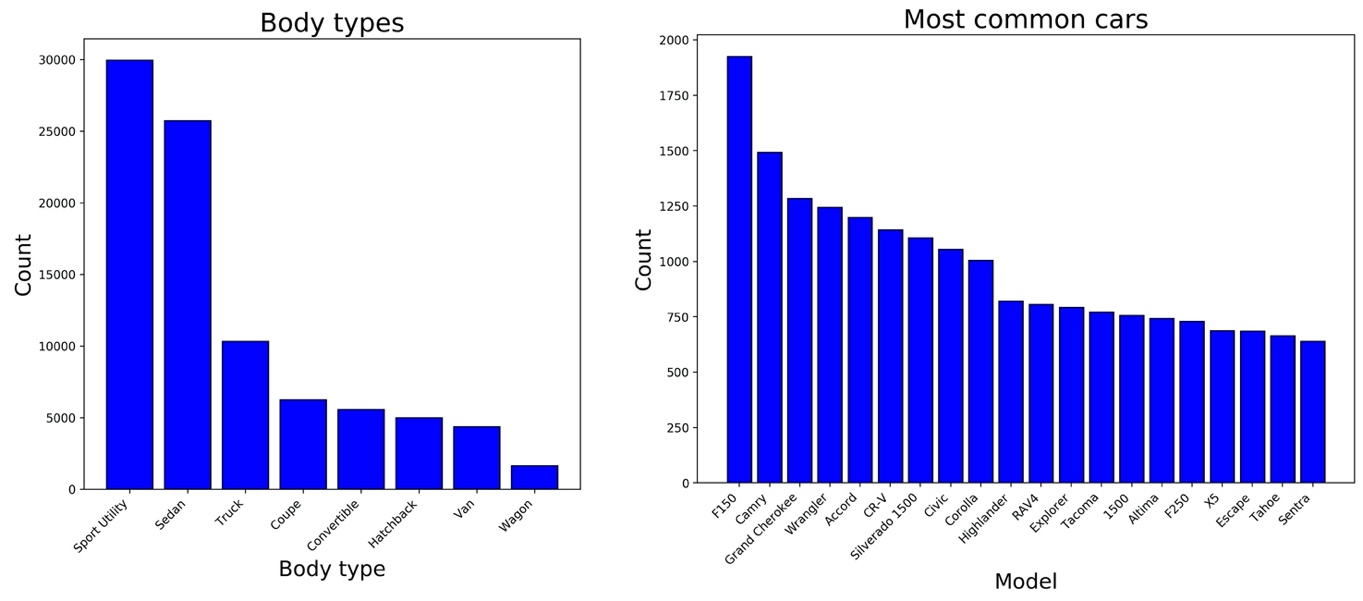


Figure 2. Histograms of (left to right) price, mileage, and miles per year across the 90,000 listings collected from Autotrader in January 2020.

1. *Depreciation analysis*
2. Price versus age and miles

With hundreds to thousands of individual listings collected for each of hundreds of car models, the evolution of price was evaluated across vehicle life cycle for each car model. For instance, plotting listings of the Honda Civic in 3D, with x, y, and z-coordinates reflecting vehicle age, mileage, and price (Figure 3).

A surface of best fit was obtained by fitting an exponential regression against age and price of the form

*P*(*t*,*m*) = (*a*/2)·[exp(-*bt*) + exp(-*cm*)] Eq. 1

Where price *P* is a function of age *t* and mileage *m*, and constants *a*, *b*, and *c* are the fitted parameters for new car price and depreciation coefficients against age and mileage, respectively. A typical scrap car commands under $300 at the junkyard (<https://www.junkcarmedics.com/blog/scrap-car-prices-per-ton/>). With a terminal value below 1% of the new price, we’ve effectively approximated it to be zero by leaving off an intercept term in Eq. 1.

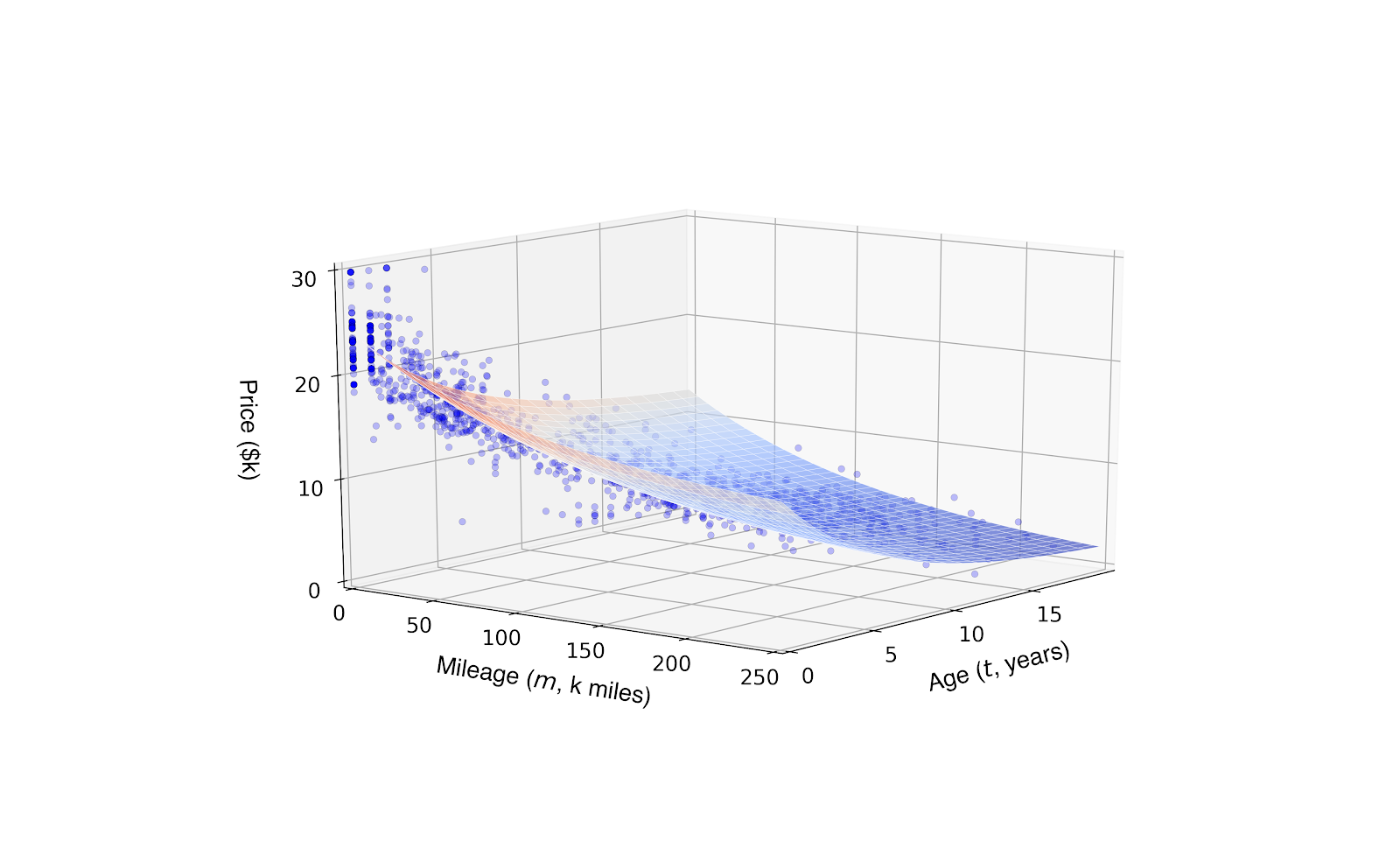


Figure 3. Scatter plot of price versus age and mileage for the Honda Civic.

1. Price versus age or miles

For most car models, the surface of best fit described by Eq. 1 explains approximately 90% of the observed variance in price.

This same listing data was examined in two dimensions: price versus age (Figure 4, left) and price versus miles (Figure 4, right).

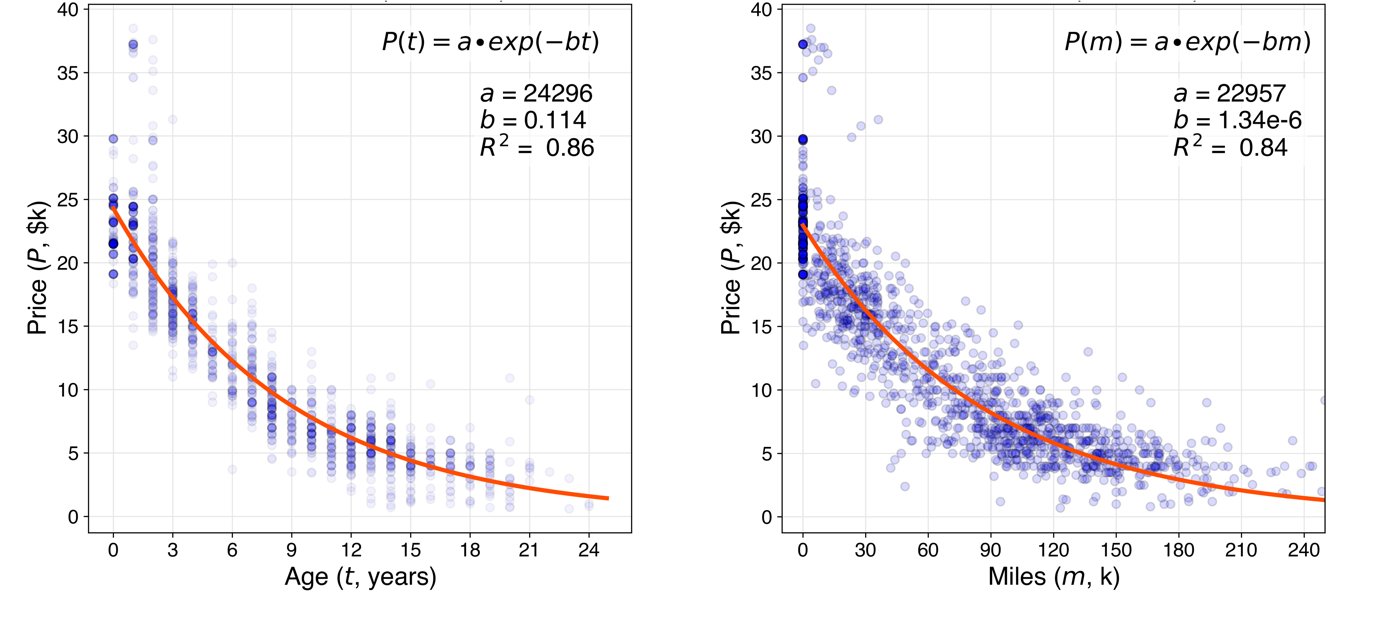


Figure 4. Scatter plot of price versus age (left) and price vs. mileage across Honda Civic listings.

Looking at the fit quality (R2 for each), age is the better univariate predictor of price than miles. The fit of price versus miles

*P*(*t*,*m*) = *a*·exp(-*bt*)

has a coefficient of exponential decay *b* that can serve as an intuitive single metric of value retention.

For most cars, fit quality is better for price versus age than price versus miles. On the other hand, some cars

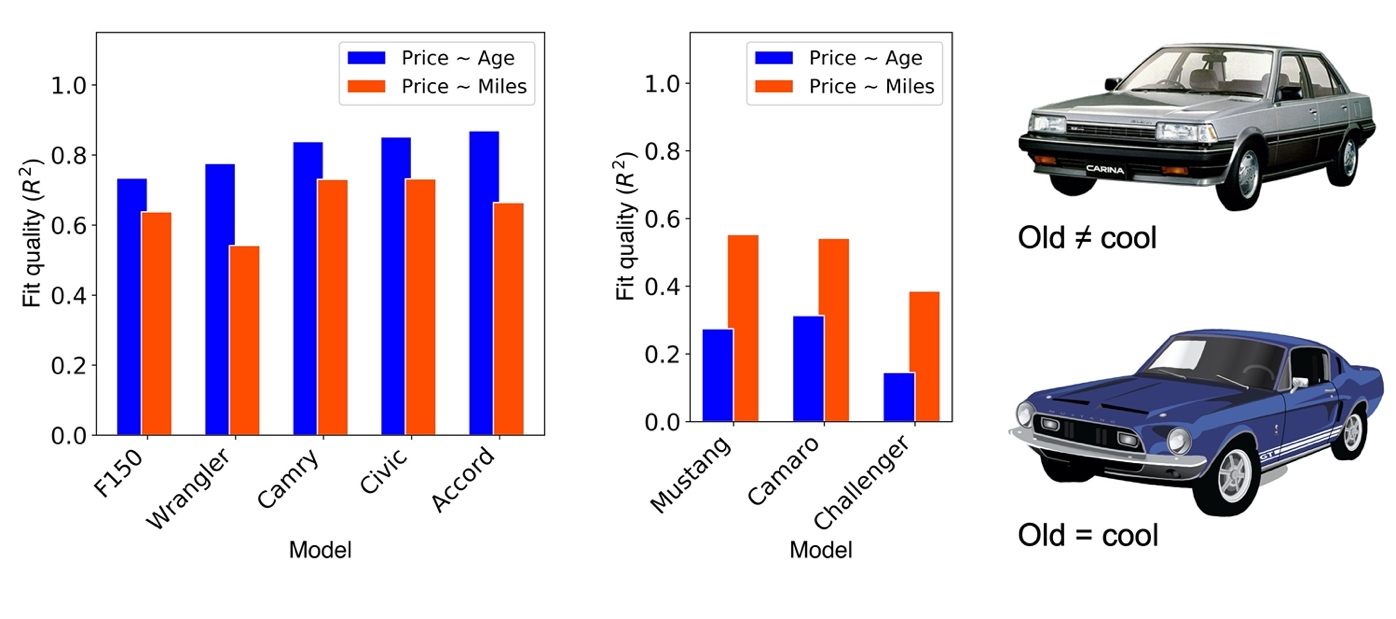


Figure 5. Scatter plot of price versus age (left) and price vs. mileage across Honda Civic listings.

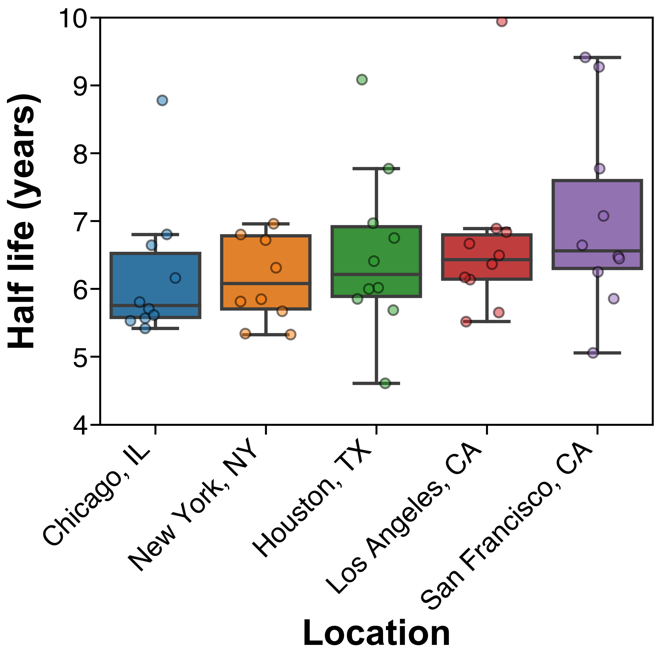
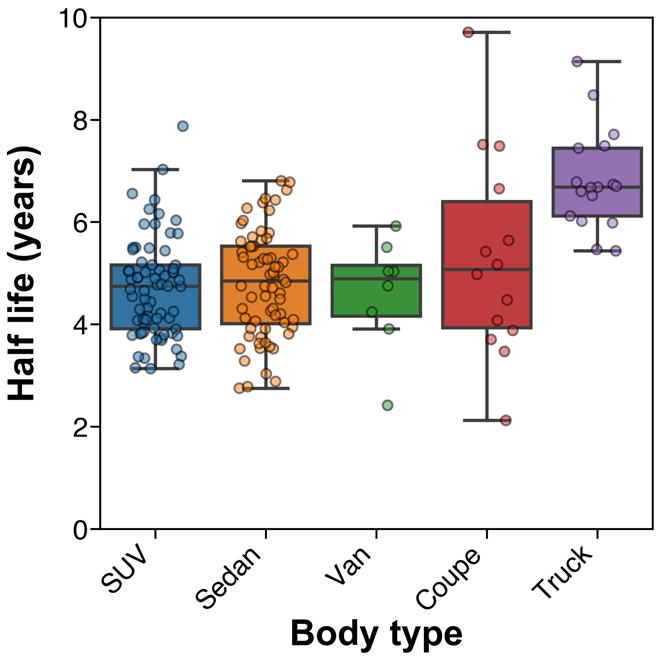


Figure 6. Depreciation rate across segments.

Figure 8. Effect of car location on depreciation rate.

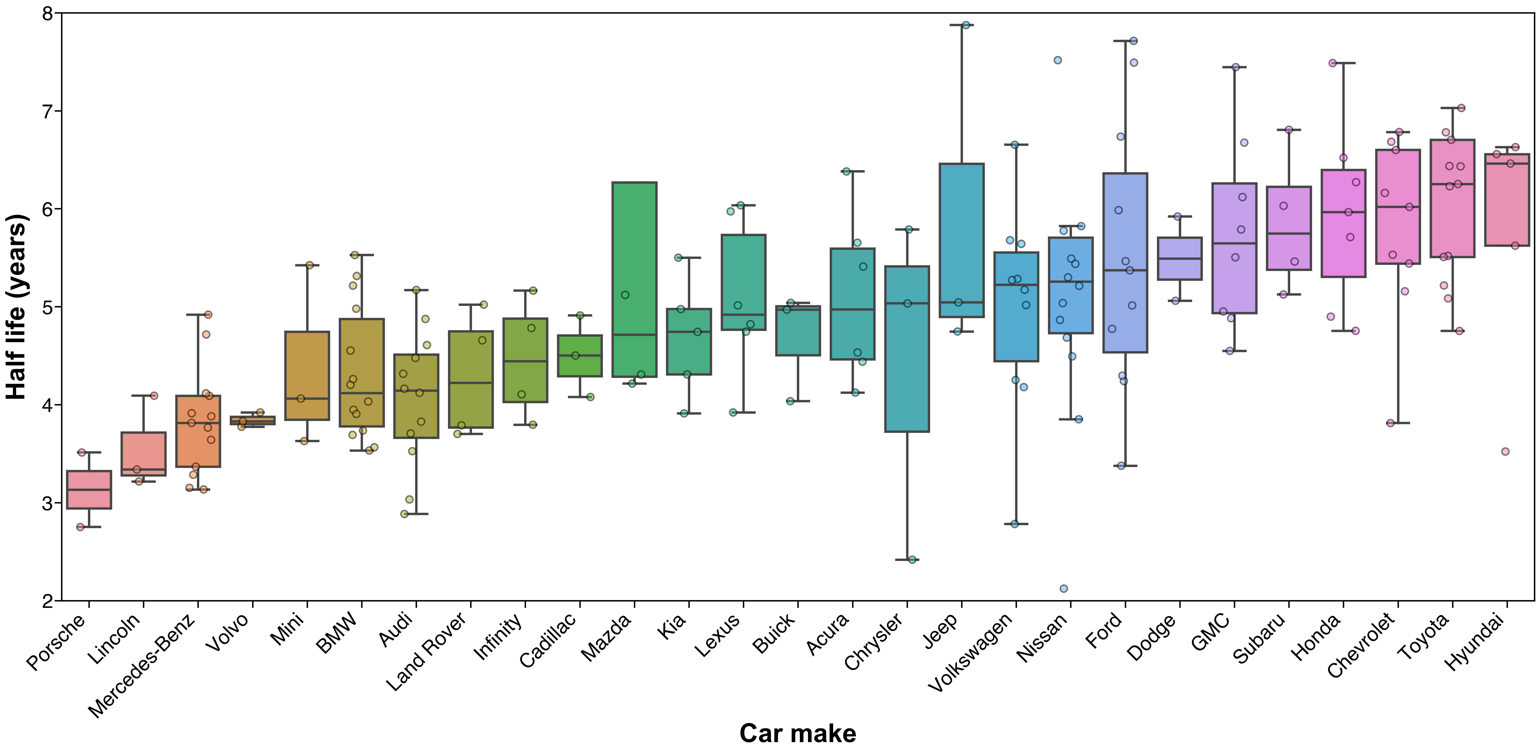


Figure 7. Depreciation rate across brands.