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

Michael Boles

February 2020

Executive summary

Introduction

Methods

Mention motivation

-car is single largest depreciating asset you’ll ever buy

-resources available to consumers are lacking

-key missing tool is depreciation estimator

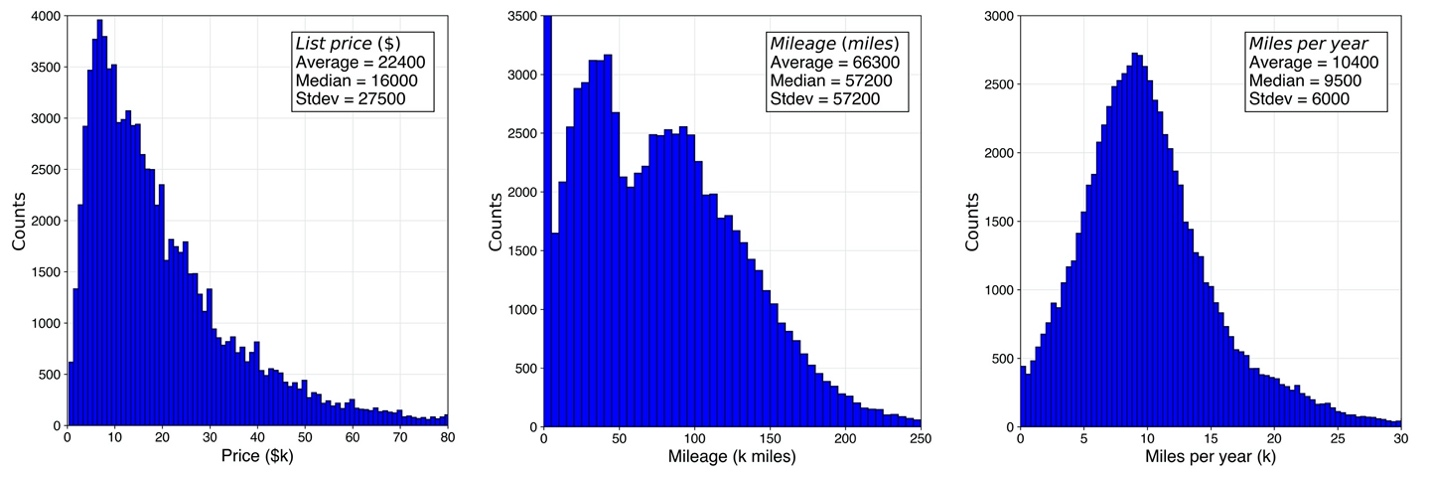
-depreciation rate is crowdsourced metric for pleasure or pain accompanying ownership of that vehicle

**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.

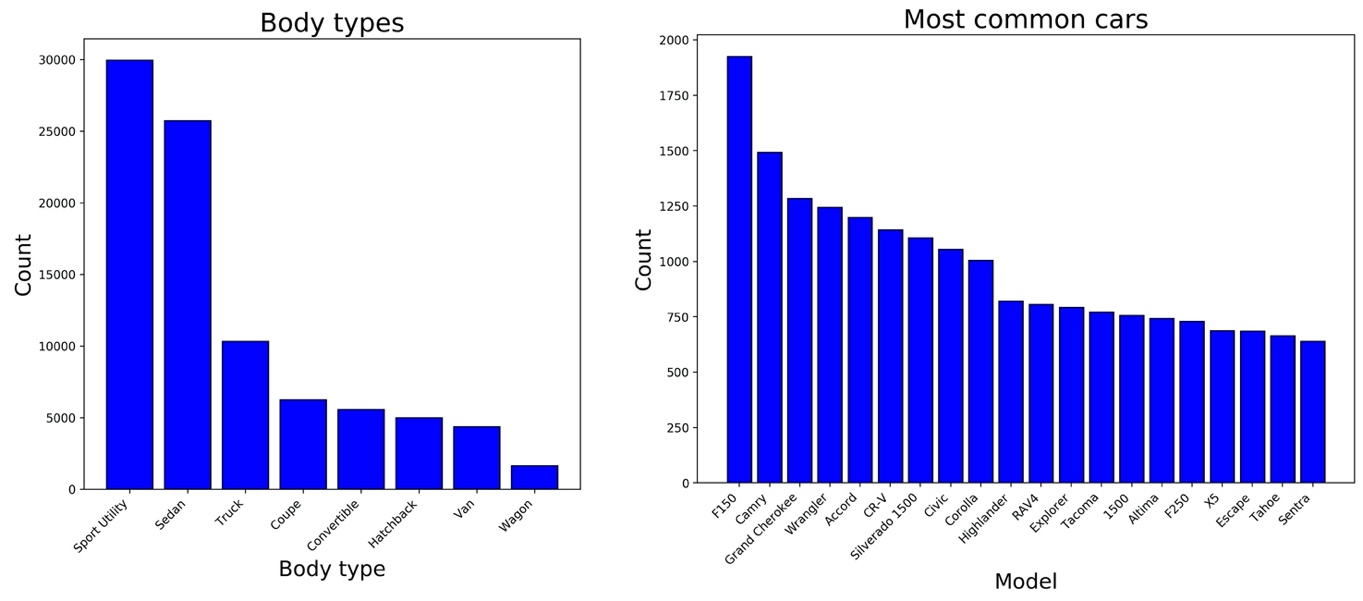
The distribution of list prices across the set (Figure 1, left) displays a long tail, with a mode of $7,000 and median price of $16,000. Reported mileage is trimodal: a large number of new cars with zero miles (20,000 listings), a second set with approximately 30,000 miles, and a third set with about 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.

Annual sport utility vehicles (SUV) sales exceeded those of sedans for the first time in 2014, (<https://www.edmunds.com/car-news/sedan-dethroned-as-most-popular-body-style-in-america.html>). This is reflected in the Autotrader listings, which show nearly 20% more SUVs than sedans (Figure 2, left). Together, SUVs and sedans make up more than half of cars on the market. On the other hand, the popularity of minivans has turned in the opposite direction. Three times more popular in the year 2000 than they are today (<https://www.freep.com/story/money/cars/2019/08/02/minivan-sales/1898974001/>), minivans 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 (by far) car in America, the Ford F150 (Figure 2, right). The top 10 most frequently encountered cars in this data set include 4 sedans (the midsize- and compact offerings of Camry, Accord, Civic, and Corolla from Toyota and Honda), 4 SUVs (Jeep’s Grand Cherokee and Wrangler alongside 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**. Bar plots showing number of listings by body type (left) and the most popular 20 models (right).

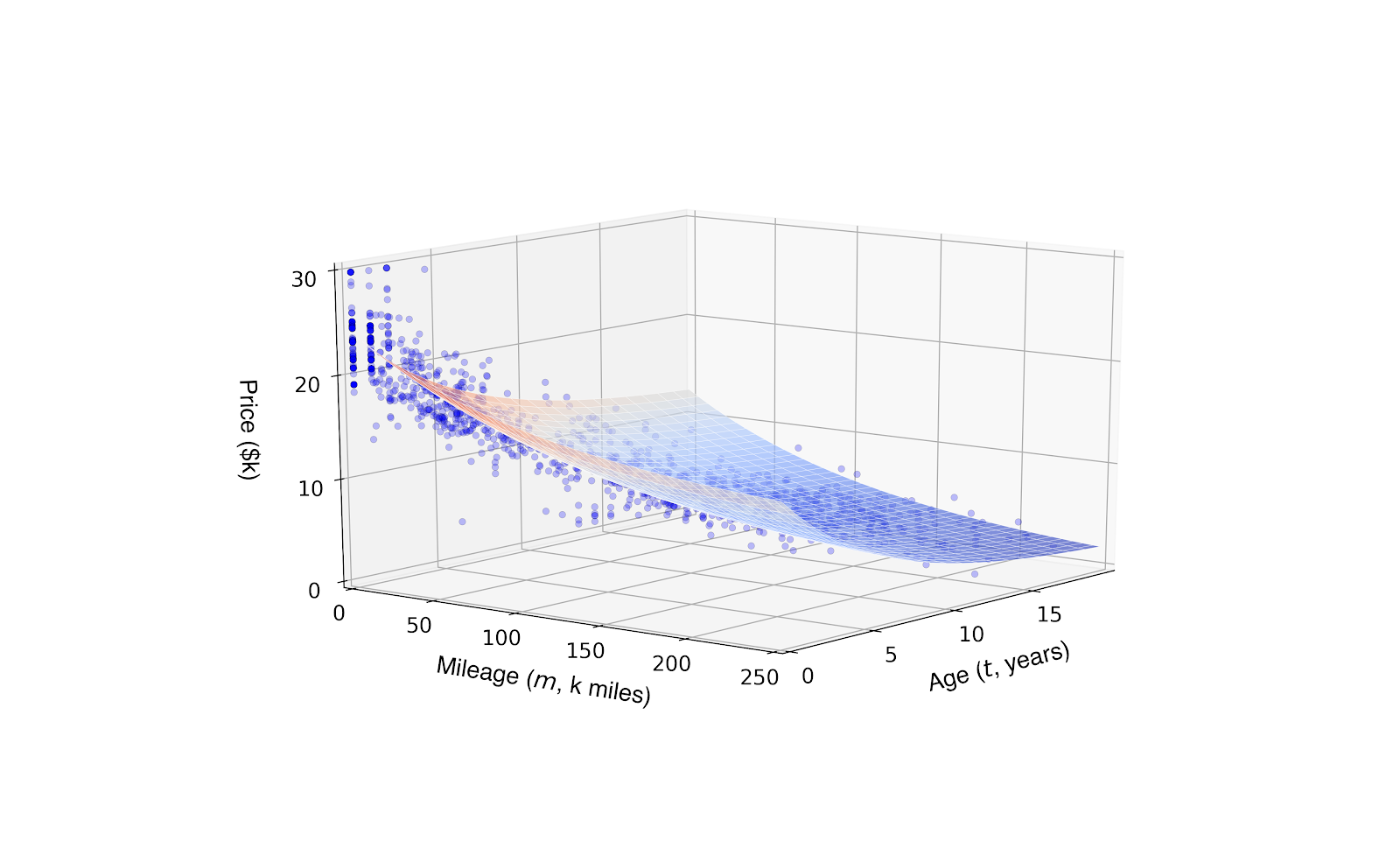
1. *Depreciation modeling*
2. Price versus age and miles

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, Figure 3 shows approximately 1500 listings of the Honda Civic plotted in 3D, with x, y, and z-coordinates reflecting vehicle age, mileage, and price.

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*)] + *d* Eq. 1

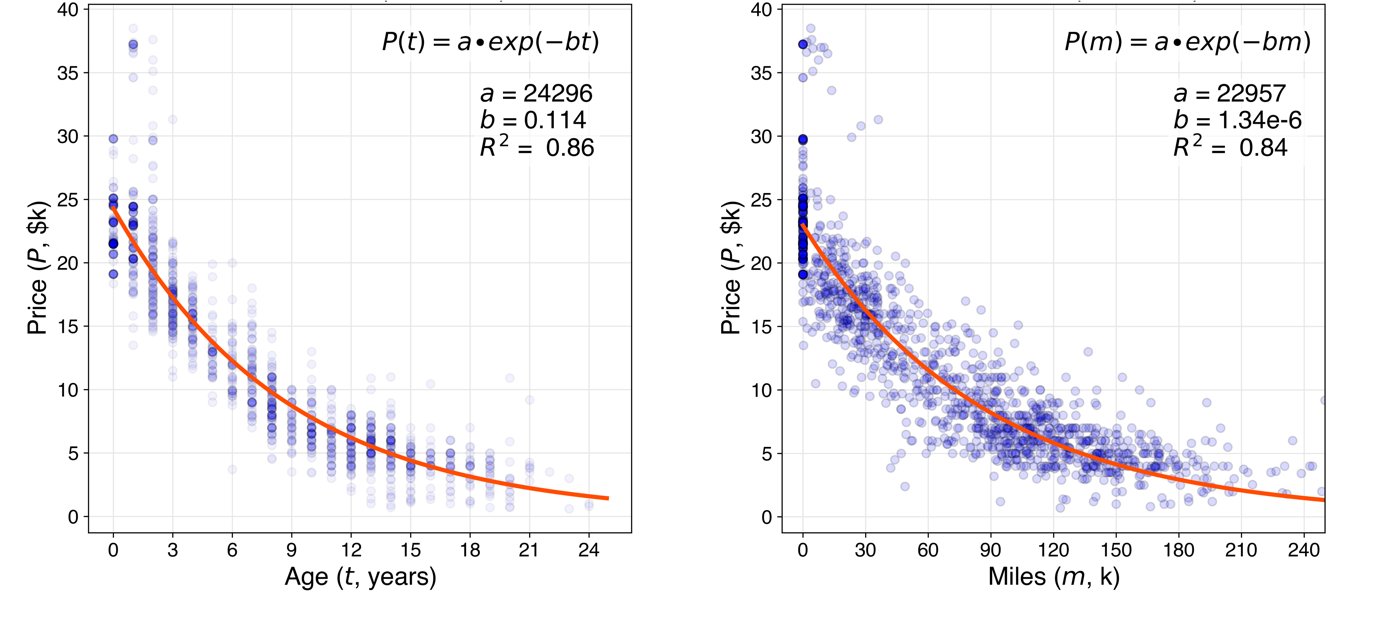
Where price *P* is a function of age *t* and mileage *m.* The new car price is captured by constant *a*, while constants *b* and *c* are the decay coefficients against age and mileage, respectively. The bias term *d* represents the terminal value of the car. Because a typical scrap car commands less than $300 at the junkyard (<https://www.junkcarmedics.com/blog/scrap-car-prices-per-ton/>), this term was left out of the fit (i.e., *d* = 0).



**Figure 3**. List price versus age and mileage for the Honda Civic (blue scatter data) and corresponding surface of best fit obtained from Equation 1.

1. Univariate analysis

For most car models, the surface of best fit described by Eq. 1 explains approximately 90% of the observed variance in price (*R*2 ≈ 0.90). When viewing the same 1500 listings in just two dimensions, either price versus age (Figure 4, left) or price versus miles (Figure 4, right), only a small sacrifice in fit quality (*R*2 ≈ 0.85) is observed.



**Figure 4**. Scatter plot of price versus age (left) and price vs. mileage (right) across Honda Civic listings. The text inset describes exponential decay functions and resulting parameters from curve fitting. Note that empirical list price data is more tightly distributed around best fit line for model employing age as predictor of price.

In each case, fitting exponential decay functions of the form

*P*(*x*) = *a*·exp(-*bx*) Eq. 2

with independent variable choice *x* (either age or mileage) allows for extraction of the typical new car price *a* and the decay coefficient *b*.

In the case of the Honda Civic, exponential fit of price vs. age (Figure 4, left) across approximately 1500 individual listings yields an estimated new car price of $24,300 and decay coefficient of 0.114 years-1.

One convenient aspect of modeling car depreciation with an exponential function is it lends itself to half-life calculation:

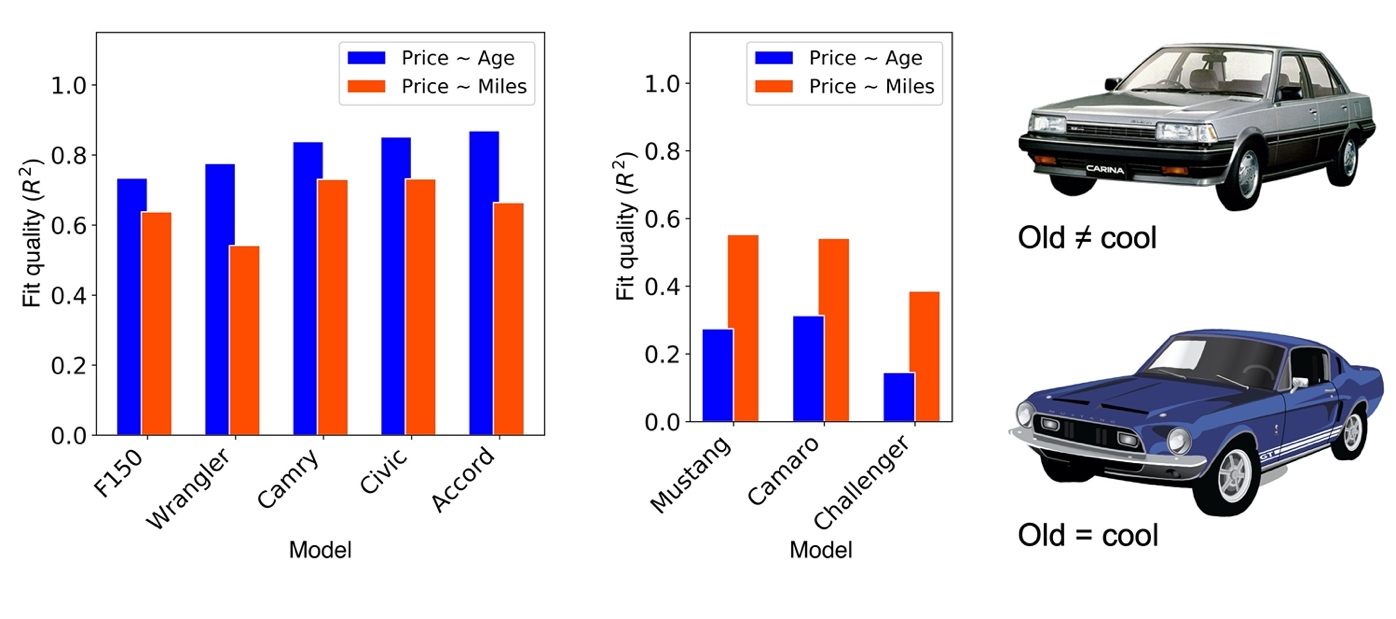
*t*1/2 = ln(2) / *b* Eq. 3

where ln(2) ≈ 0.693 is the natural logarithm of 2 and *b* is the decay coefficient determined by exponential fit.

1. Model selection

While the model constructed with both vehicle age and mileage (Eq. 1) offers a slightly better fit quality than a univariate model (Eq. 2), the latter yields just one decay coefficient (*b*) instead of two (*b* and *c*), and thus offers a convenient single metric, or figure of merit, describing vehicle value retention.

Fit quality (*R*2) was then used to decide between age and mileage as the best univariate predictor of price. Across all five most frequently encountered listings, age seems to do better than miles (Figure 5 left, blue versus orange bars). There is, however, a small subset of models that show the opposite trend. For cars such as the Ford Mustang, Chevy Camaro, and Dodge Challenger, age turns out to be a poor predictor of list price (Figure 5 middle). In these cases, there is a huge variation in price for the oldest models: while some may be on their way to the junkyard, others are trading hands for many multiples of their original MSRP. For cars with this sort of vintage appeal, it is difficult to estimate value and, as a result, depreciation rates.

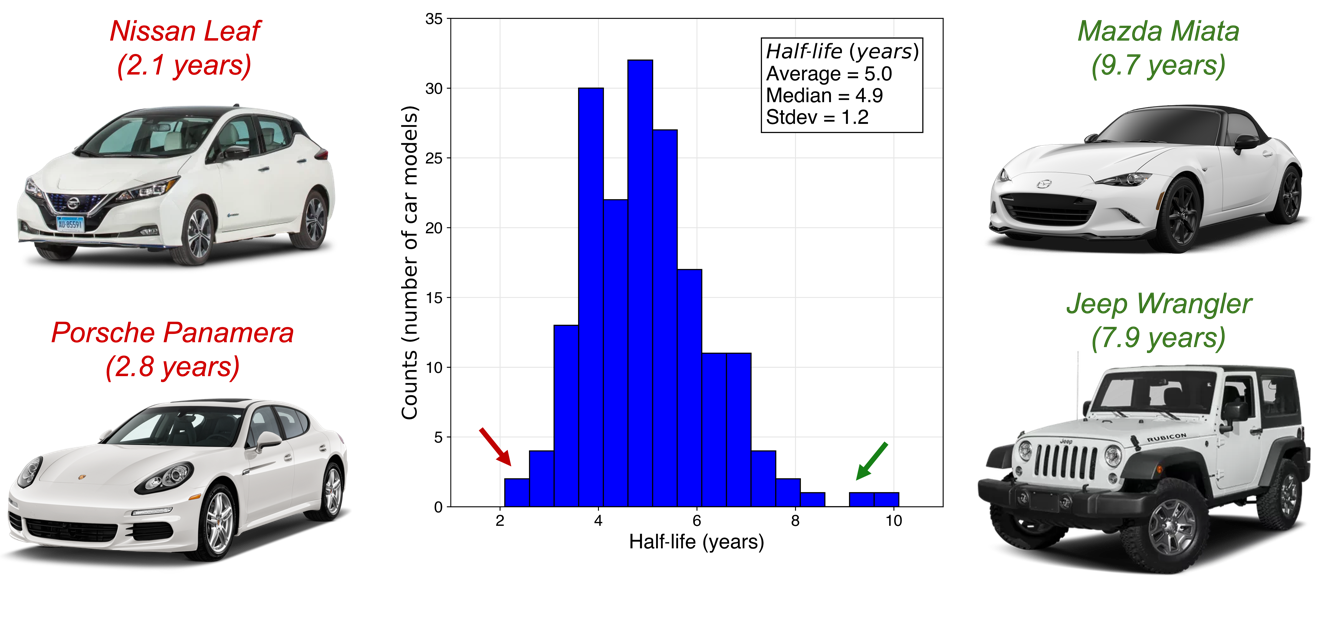


**Figure 5**. Scatter plot of price versus age (left) and price vs. mileage across Honda Civic listings.

1. *Results*
2. Outliers: best and worst in value retention

For each of several hundred make/model combinations present in the Autotrader dataset, listing data was fit according to Eq. 2 using vehicle age as the independent variable (Figure 4, left). Empirical depreciation curves for cars with a small number of observations (< 50 listings) or anomalous pricing (e.g., vintage appeal) typically yielded poor fit quality. To avoid drawing spurious conclusions from poorly understood pricing, a fit quality filter (*R*2 > 0.67) was applied.

Across 178 models with well-characterized pricing, the observed range of vehicle half-lives spans 2 to 10 years, with most cars typically experiencing a 5-year half-life (Figure 6, center). Among the fastest-depreciating cars are electric vehicles (EVs) and foreign luxury cars, while those that retain their value the longest are simple, mass-produced, and arguably iconic cars (Figure 6, left and right).



**Figure 6**. Histogram (center) displaying vehicle half-life estimates across 178 models. Each count was obtained by extracting the coefficient of exponential decay as shown in Figure 4 for hundreds (or thousands) of individual Autotrader listings for a particular model. Prototypical examples of quickly-depreciating and slowly-depreciating vehicles are shown at left and right, respectively.

Perhaps unsurprisingly, foreign luxury cars are close to the bottom of the list when it comes to value retention. One factor is the high maintenance costs that accompany ownership of such cars (<https://www.businessinsider.com/10-cars-lose-the-most-value-last-5-years-2019-10>). In addition, the inevitable blemishes and dated appearance that used cars bring may be particularly undesirable to used luxury car buyers, who are almost certainly more image-conscious than others.

This analysis also indicates that the depreciation rate for EVs appears to be surprisingly harsh. Some of this effect may be artificial, since the $7500 Federal tax credit on the purchase of new EVs, which will be phased out over 2020 (<https://cleanvehiclerebate.org/eng/ev/incentives/state-and-federal>), is not available to used EV buyers and is thus immediately dropped off the used car sticker price. Even so, consumers seem to perceive EV technology as one that is advancing rapidly, and that used EVs may be by definition outmoded. In addition, range anxiety might be more pronounced in the used market: EV batteries begin to lose cruising range at 100,000 miles and may need replacement (at a cost upwards of $15,000) when the odometer approaches 200,000 miles (https://www.myev.com/research/ev-101/how-long-should-an-electric-cars-battery-last). Interestingly enough, it seems as though the Tesla brand is an exception to the EV depreciation rule (<https://www.greencarreports.com/news/1123583_beyond-tesla-electric-cars-lose-value-faster-than-other-vehicles>), with possible explanations for this being more robust battery life, drivetrain, and sensors (<https://www.tesloop.com/blog/2019/2/6/tesla-and-the-electrifying-economics-of-depreciation>). In this analysis, half-life estimates of ~5 years for the Tesla Model S and Model X place them in the middle of the pack, although fit quality (*R*2 < 0.5) indicates this figure should be interpreted with some caution.

At the other end of the spectrum, the Mazda Miata and Jeep Wrangler (<https://www.roadandtrack.com/new-cars/g16345197/cars-with-slow-depreciation-that-hold-value/?slide=1>) are some of the best performers when it comes to value retention. These cars, and others that appear to retain their value well, are affordable, mass-produced, and have been in production for several decades.

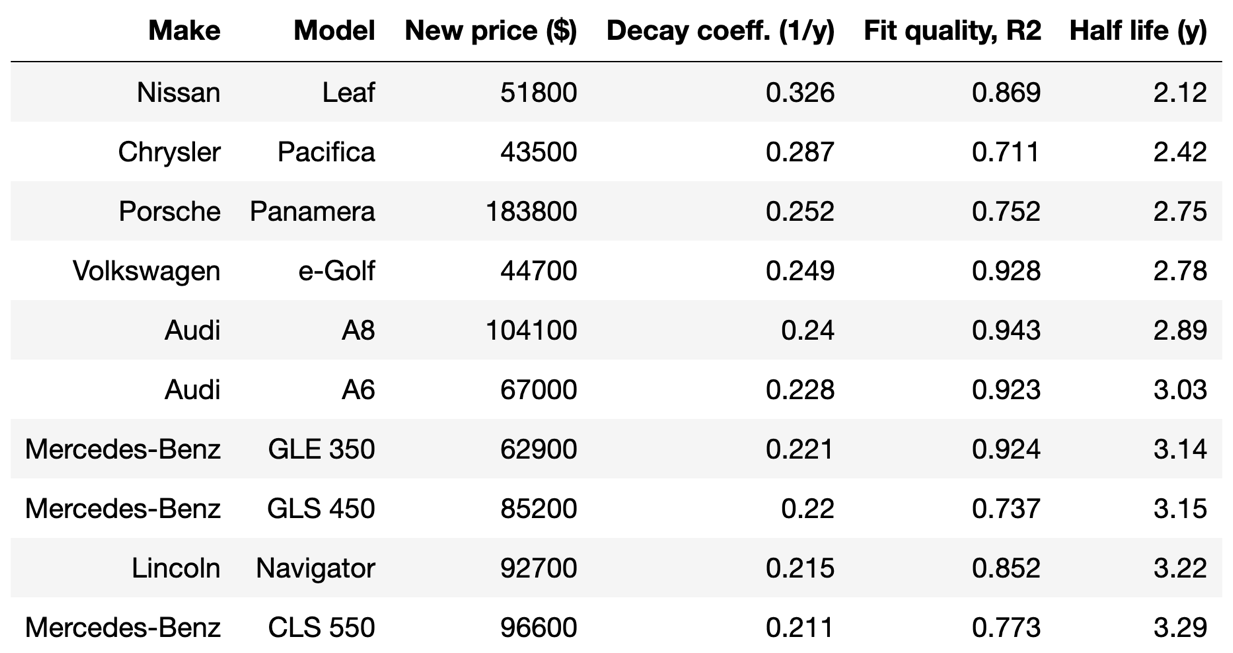
Cars that perform the best in terms of value retention are shown in Table 1. This table displays, for a given make and model, the new price (extracted as fit parameter *a*, in dollars), the rate of decay (fit parameter *b*, with units years-1), the resulting half-life (calculated as ln(2)/*b*, in years), and the fit quality (*R*2 value, or fraction of observed variance that is explained by the model). Six of the top 10 cars for value retention belong to Japanese brands (Toyota, Nissan, Mazda, Honda), while the other four are American (Jeep, Ford, GMC). These cars typically lose half their value every 7 to 10 years, and are disproportionately entry-level or utility-focused.

**Table 1**. Top ten cars for value retention.



At the other end of the spectrum, cars that appear to lose value the quickest are shown in Table 2. Seven of the bottom ten cars for value retention are luxury offerings, and most of them are German (Porsche, VW, Audi, Mercedes). On average, these cars are twice as expensive as those appearing in Table 1, and two of the bottom four are EVs (the Nissan Leaf and VW e-Golf).

**Table 2**. Bottom ten cars for value retention.

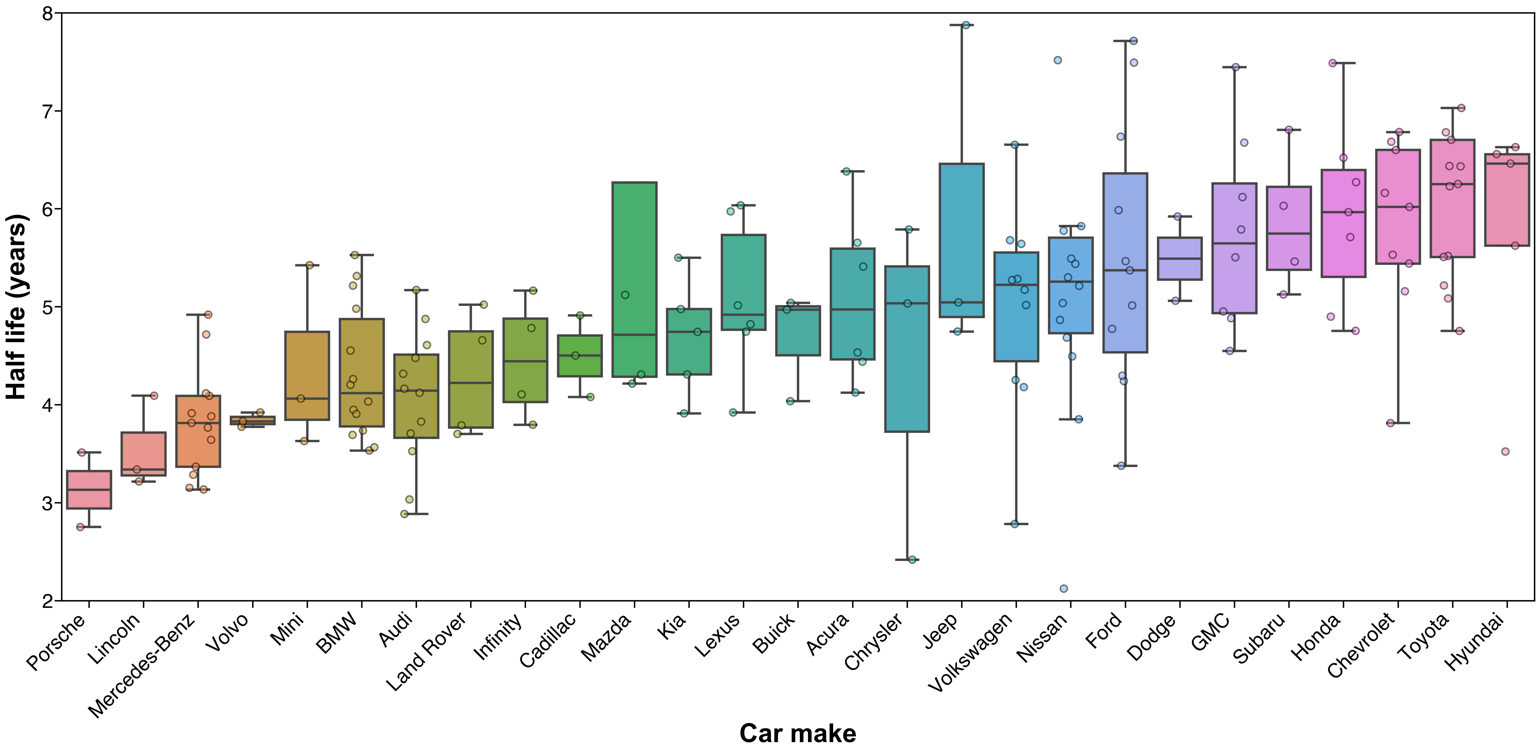


In striking contrast to those displayed in the previous table, these cars lose half their value every 2 to 3 years. To put this into perspective: a $50,000 car with 7-year half-life can be sold for $25,000 after 7 years, while one with a 3-year half-life is worth just under $10,000 after the same time interval. Put another way: choosing a slowly-depreciating car over a similarly-priced, quickly-depreciating one is, upon selling after 7 years, equivalent to receiving into your bank account a transfer of 30% of the initial value of the vehicle.

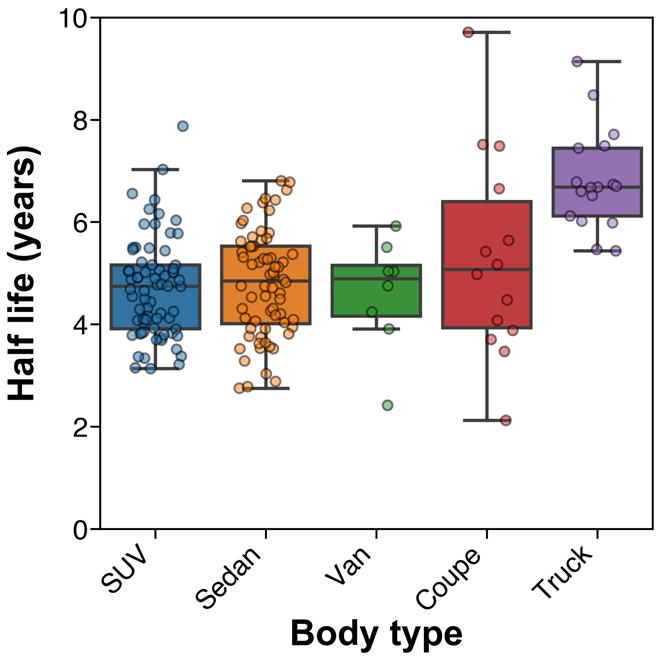
1. Depreciation trends

Perhaps more interesting than comparing empirical depreciation rates across individual car models is looking at aggregated data across other axes such as brand, segment, location, etc.

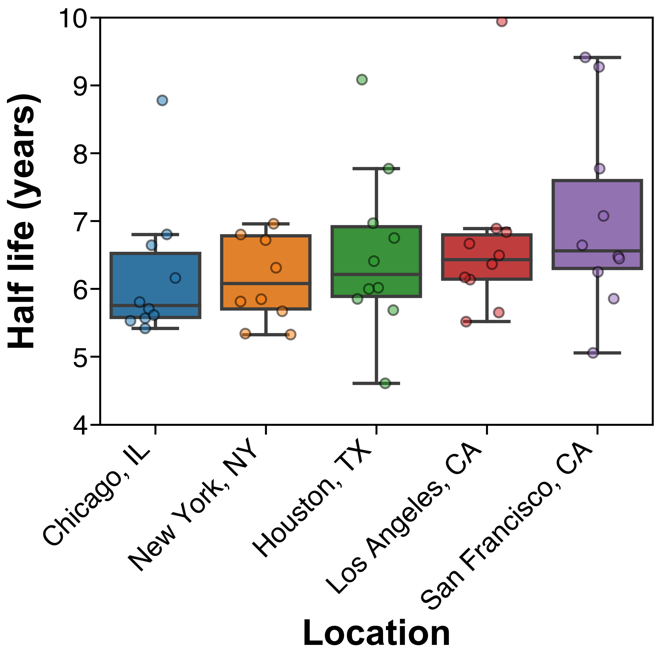
To this end, each of the 180 car models with well-characterized pricing (*R*2 > 0.67 for exponential fit of price versus age) was grouped by car make, and the corresponding half-life was visualized by brand in a box plot (Figure 7). Across all brands, average half-lives span 3 to 6 years. Strikingly, the 10 brands that appear to lose their value the quickest are all in the luxury segment. Among luxury offerings, Lexus and Acura stand out with their solidly average half-lives of about 5 years. At the other end of the spectrum, 4 of the best 5 car brands for value retention are either Japanese or Korean. Hyundai and Toyota are at the top of this list, with all models (except for their luxury models Genesis and Avalon) showing half-lives above 5 years.



**Figure 7**. Depreciation rate across brands. For each make, the colored box denotes the interquartile range (25th to 75th percentile), the line within it denotes the median value, and the whiskers encompass the remainder of the distribution minus outliers. Each scatter point represents, for a particular model, the decay coefficient resulting from exponential fit of hundreds or thousands of individual Autotrader listings.



**Figure 8**. Depreciation rate across segments.



**Figure 9**. Effect of listing location on depreciation rate, binned by vehicle segment.