Data Analysis: A Practical Introduction for Absolute Beginners

Lab 2: Real World Data

Learning Objectives

- Look at the basic anatomy of a data set with several variables.
- Sort that data according to one particular variable in Excel.

Data Set

Mod2Labs.csv

What You'll Need

To complete the lab, you will need the online version of Microsoft Excel.

Overview

In this lab, we'll take a closer look at the car data we looked at in the videos. We'll also see how to sort the data based on a specific variable/criteria, such as gas mileage.

Exercise 1: Car Data

1. Open the data set in Excel. You should see several different variable columns for a group of car models. Here's a snapshot of the data:

	A	B	C	D	E	F	G	H	I	J	K	L
1	model	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
2	Knight X	21	6	160	110	3.9	2.62	16.46	0	1	4	4
3	Knight X Wagon	21	6	160	110	3.9	2.875	17.02	0	1	4	4
4	Hercules 100	22.8	4	108	93	3.85	2.32	18.61	1	1	4	1
5	Wasp 4WD	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
6	Wasp Supersport	18.7	8	360	175	3.15	3.44	17.02	0	0	3	2
7	El Pasion	18.1	6	225	105	2.76	3.46	20.22	1	0	3	1
8	Road Devil	14.3	8	360	245	3.21	3.57	15.84	0	0	3	4
9	Anansi 100	24.4	4	146.7	62	3.69	3.19	20	1	0	4	2
10	Anansi 200	22.8	4	140.8	95	3.92	3.15	22.9	1	0	4	2
11	Anansi 200x	19.2	6	167.6	123	3.92	3.44	18.3	1	0	4	4

2. Identify the variables. In this case, there are 12 different vertical **columns** . Here's what they each represent:

   model = name of car model

   mpg = gas mileage, in miles per (US) gallon

   cyl = number of cylinders

   disp = displacement, in cubic inches

   hp = gross horsepower **drat** = rear axle ratio

   wt = weight, in thousands of pounds (1000 lb)

   qsec = 1/4 mile time

   vs = engine (0 = V-shaped, 1 = straight)

   am = transmission (0 = automatic, 1 = manual)

   gear = number of forward gears

   carb = number of carburetors

   Those are your variables, since they vary from car to car.

3. Identify the observations. Since each horizontal **row** corresponds to a single entry from each of the variable columns, each row must represent an individual car model. For example, take a look at the first car on the list: the Knight X gets 21 miles per gallon, has 6 cylinders, 110 horsepower, etc.

	A	B	C	D	E	F	G	H	I	J	K	L
1	model	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
2	Knight X	21	6	160	110	3.9	2.62	16.46	0	1	4	4
3	Knight X Wagon	21	6	160	110	3.9	2.875	17.02	0	1	4	4
4	Hercules 100	22.8	4	108	93	3.85	2.32	18.61	1	1	4	1
5	Wasp 4WD	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
6	Wasp Supersport	18.7	8	360	175	3.15	3.44	17.02	0	0	3	2

4. You can also answer simple questions about the data. For instance, how much does the Clydesdale Turbo weigh? To find out, track down "Clydesdale Turbo" in the model column, then look over at column G, the weight column.

	A	B	C	D	E	F	G	H	I	J	K	L
13	Anansi 350 Deluxe	16.4	8	275.8	180	3.07	4.07	17.4	0	0	3	3
14	Anansi 400	17.3	8	275.8	180	3.07	3.73	17.6	0	0	3	3
15	Anansi 400x	15.2	8	275.8	180	3.07	3.78	18	0	0	3	3
16	Clydesdale Turbo	10.4	8	472	205	2.93	5.25	17.98	0	0	3	4
17	Seaward	10.4	8	460	215	3	5.424	17.82	0	0	3	4
18	Commonwealth 360	14.7	8	440	230	3.23	5.345	17.42	0	0	3	4

Remember, the weight variable ("wt") is given in *thousands* of pounds, so multiply 5.25 by 1,000 to find the car's weight.

5.25 × 1,000 = 5,250

The Clydesdale Turbo weighs in at 5,250 pounds.

5. What kind of horsepower does the Apocalypse 100 get? To find out, find that car model in the column on the far left, then look at column E to see the horsepower ("hp").

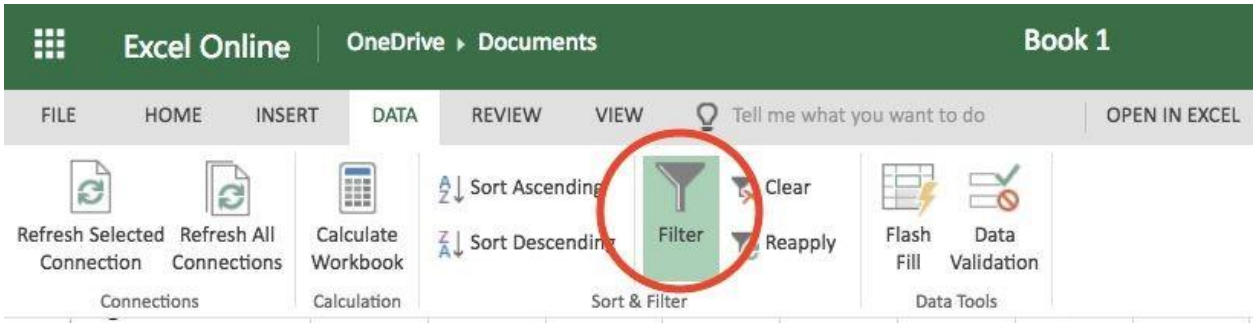
	A	B	C	D	E	F	G	H	I	J	K	L
24	Tiara	15.2	8	304	150	3.15	3.435	17.3	0	0	3	2
25	Osprey Z-7	13.3	8	350	245	3.73	3.84	15.41	0	0	3	4
26	Hollinger Towncar	10.2	8	400	175	3.08	3.845	17.05	0	0	3	2
27	Apocalypse 100	27.3	4	79	66	4.08	1.935	18.9	1	1	4	1
28	Prince of Thieves	26	4	120.3	91	4.43	2.14	16.7	0	1	5	2
29	Excalibur Phantom	30.4	4	95.1	113	3.77	1.513	16.9	1	1	5	2

Nice. That particular model gets 66 horsepower. Actually… that's not so nice.

Exercise 2: Sorting by MPG

Now we'll rearrange the data so we can easily see which car models get the best and worst gas mileage.

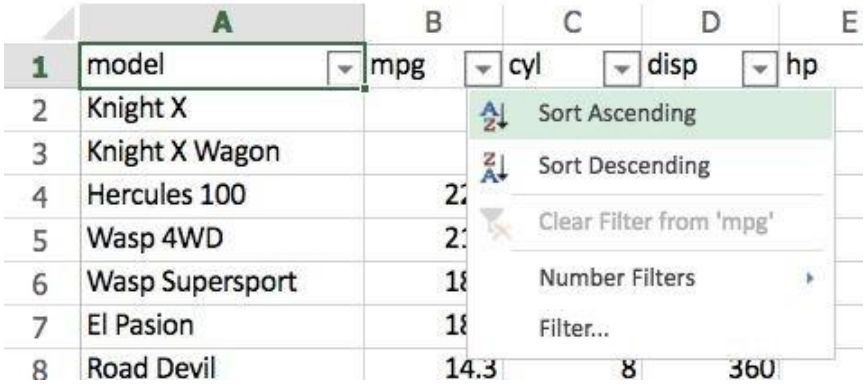
1. With the data set open, click anywhere in the spreadsheet, go to the Data tab in the ribbon, and click Filter (in the Sort & Filter tab).



2. You should now see a little dropdown arrow show up next to each column title, like so:

	A	B	C	D	E	F	G	H	I	J	K	L
1	model	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
2	Knight X	21	6	160	110	3.9	2.62	16.46	0	1	4	4
3	Knight X Wagon	21	6	160	110	3.9	2.875	17.02	0	1	4	4
4	Hercules 100	22.8	4	108	93	3.85	2.32	18.61	1	1	4	1
5	Wasp 4WD	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
6	Wasp Supersport	18.7	8	360	175	3.15	3.44	17.02	0	0	3	2

3. That arrow allows you to sort all the data by any single variable. We want to arrange the data based on gas mileage, so click the arrow next to “mpg” and select Sort Ascending.



4. Now the entire spreadsheet should rearrange itself, showing the car with the lowest gas mileage at the top and ascending downward. The great thing about this feature is that it shuffles *all* the data based on one particular column/variable.

	A	B	C	D	E	F	G	H	I	J	K	L
1	model	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
2	Clydesdale Turbo	10.4	8	472	205	2.93	5.25	17.98	0	0	3	4
3	Seaward	10.4	8	460	215	3	5.424	17.82	0	0	3	4
4	Osprey Z-7	13.3	8	350	245	3.73	3.84	15.41	0	0	3	4
5	Road Devil	14.3	8	360	245	3.21	3.57	15.84	0	0	3	4
6	Commonwealth 360	14.7	8	440	230	3.23	5.345	17.42	0	0	3	4
7	Stegosaurus Diesel	15	8	301	335	3.54	3.57	14.6	0	1	5	8
8	Anansi 400x	15.2	8	275.8	180	3.07	3.78	18	0	0	3	3
9	Tiara	15.2	8	304	150	3.15	3.435	17.3	0	0	3	2
10	Reisender Bluebird	15.5	8	318	150	2.76	3.52	16.87	0	0	3	2

5. With the data sorted like this, you can more easily see, for example, that the Clydesdale Turbo gets the lowest gas mileage out of any car on the list (10.4 miles per gallon). At the bottom of the list, you can see that the Empire Baroness gets the best gas mileage (33.9 miles per gallon).

   We’ll go with the Baroness when we buy a new car.