CSV Data as Truenumbers: Solar System Data

The Data

We open the CSV file in **pandas** to get a data frame of solar system information. We select columns of interest for this brief and sort on a column that brings rows of interest to the top. We then display the first 20 rows.

```
import pandas as pd
df = pd.read_csv ("C:/Users/allen/Documents/Jupyter/sol_data.csv",
usecols=['eName','isPlanet','mass_kg','orbit_type','orbits'])
sd = df.sort_values(by=['isPlanet'], ascending=False)
sd.head(15)
```

	eName	isPlanet	mass_kg	orbit_type	orbits
199	Uranus	True	8.680000e+25	Primary	NaN
244	Venus	True	4.870000e+24	Primary	NaN
241	Saturn	True	5.680000e+26	Primary	NaN
240	Mercury	True	3.300000e+23	Primary	NaN
179	1 Ceres	True	9.390000e+20	Primary	NaN
239	Mars	True	6.420000e+23	Primary	NaN
208	Pluto	True	1.300000e+22	Primary	NaN
238	Jupiter	True	1.900000e+27	Primary	NaN
236	136472 Makemake	True	4.400000e+21	Primary	NaN
243	Earth	True	5.970000e+24	Primary	NaN
219	Neptune	True	1.020000e+26	Primary	NaN
187	136199 Eris	True	1.660000e+22	Primary	NaN
222	136108 Haumea	True	4.010000e+21	Primary	NaN
171	Hiʻiaka	False	2.400000e+16	Secondary	136108 Haumea
173	Dia	False	9.000000e+13	Secondary	Jupiter

This dataset was sourced from kaggle.com. The data description says that some of the data was retrieved from research papers and some of it was calculated. In this brief, we're trying to give a sense of the thought process involved in designing TNs with a few example of what they might be. We don't show the process of developing code to generate those TNs. Here are Kaggle's descriptions of the columns we selected.

column name	type	description
eName	string	the name of the object
isPlanet	boolean	is the object a planet (this includes the five dwarf planets)
mass_kg	integer	total estimated mass of object in kg
volume	integer	approximate volume in km^3
orbit_type	class	either primary; orbites the Sun, or secondary; orbits a planet

column name	type	description
orbits	class	the planet that the body orbits. If it does not orbit a planet then it is NA

Discussion of the data

As with most CSV data, some of the columns help identify the subject of the row, where others give values for *properties* of that subject. This is not always a clear division. In this data, Each row appears to be about a prticular planet or other body in the Solar System, named in the **eName** column. The **mass_kg** column is clearly the measure of a *property* of the body so we expect at least one TN like

Mars has mass = 2.4e16 kg

We'll want the other columns to help make the subject phrase more informative than just the **eName**. Or, those columns might be used as values of a property like **mass** was. A little research shows us that classification and naming of solar system bodies is not simple. Planets orbit the sun by definition, and all but the major planets we learned in grade-school are considered *minor planets*. The International Astronomical Union (IAU) prefers the term *small Solar System body* but we'll stick with *minor planet*. Five of them have been designated *dwarf planets* because they share hydrostatic properties with the major planets. The data at hand does give us information that would let us generate proper subject phrases, but in kind of an oblique way.

isPlanet, if **True**, tells us that the object is some kind of planet. **orbit_type** does too, by calling a solar orbit *primary* and *secondary* is for *satellites* of some other body, usually a planet, named in the **orbits** column. For some reason, that column is blank for planets instead of containing *sun*. The **eName** signals the major planets by having no number in front of the name, but there appears to be no indication of *dwarf planet* status, so we either ignore that distinction or just use the knowledge that there are only 5 of them in building our subject phrases.

Sketching our TNs

Without getting into the task of programming a conversion, we can write down some examples from the rows above.

dwarf planet Ceres has mass = 9.39 x 10^20 kg
Hydra satellite of dwarf planet Pluto has mass = 2.4 x 10^16 kg
dwarf planet Eris (136199-Eris, trans-neptunian) has mass = 1.66 x 10^22 kg

We don't know what purpose the CSV served originally, but it was certainly not for reference material. In general, we want TNs to combine data with information so that separate descriptions are not needed when the TNs are viewed by the community of interest for that data. This more often than not requires augmenting the data source as we've done here.

The post and index test

Posting these three TNs to a numberspace using the TN Web Dashboard does two things. First, it lets us fix any grammar or illegal character problems that prevent them from posting. They were fine in this case. Second, we can look at the subject index for those numbers to see if it gives a good overview of the topics the Tns are about. A portion of that index is shown below and does a pretty good job.

Table   **Term Index**

Viewing Path Terms by

C ▾
1 related terms

  • **Ceres** ▸
    1 mention

D ▾
1 related terms

  • **dwarf** ▾
    3 mentions
    mentioned in
        ○ dwarf planet Ceres ⋮
        ○ dwarf planet Eris ⋮
        ○ Hydra satellite of
           dwarf planet Pluto   ⋮

E ▾
1 relat

  •

P ▾
2 related terms

  • **planet** ▸
    3 mentions

  • **Pluto** ▸
    1 mention

S ▾
1 related terms

  • **satellite** ▸
    1 mention

In [ ]: