CSC343 Phase 1 Submission

- Student names: Jason Li, Guanyu Song

Our Data

Data to be sourced from: https://api.le-systeme-solaire.net/en/

Most of the data in the dataset will be useful for our investigations. The information catagorizing the bodies into different sections such as moons and planets will help us focus our investigations. Since two of our questions are about the physical properties of the bodies, the data on the volume, radius, mass, and gravity will be useful in asnwering our questions. Finally, any information regarding astronomers will also helpful since our final questions is regarding astronomers.

For the investigation, we will both have to review how to analyze data to create meaningful mathematical relationships. We may also have to learn some basic data analysis methods.

As for cleaning up the data, we also need to do some quick research either by hand or with a web scrapper to get the nationalities of the people in the data base. This should be do able since most, if not all of astronomers have an historical or modern online presence. We will also choose to not add or ignore the irrevent data in the dataset. However, we suspect this won't be a major issue since we will be using the majority of the data in the dataset for our investigation.

Exploration Questions

1. Which nation has contributed most to the discovery of celestial bodies among different historical periods? Are there spikes in discoveries with the advent of new technologies?

2. Is there a relationship between volume, mass, gravity, and escape velocity? Can we derive the physics formulas ourselves using the data?

3. What is the relationship between number of moons and planet volume, mass, and gravity?

Schema

Planet(<u>bid</u>, name, type, mass, volume, moon_num, escape, gravity)

A tuple in this relation represents a planet. *bid* is an unique body identification number. *name* is the name of the planet, *mass* is the mass of the planet, and *moon_num* is the number of natural satellites the planet has. *escape* is the speed needed for a rocket to escape the planet's gravitational field. *mass* is the mass of the planet.

SmallMoon(<u>bid</u>,name,mass)

A tuple in this relation represents a moon of a planet not large enough to be spherical. *bid* is a unique body identification number for this moon. *name* and *mass* are the name and mass of the moon respectivly.

LargeMoon(<u>bid</u>, name, mass, volume, escape, gravity)
A tuple in this relation represents a moon of a planet large enough to be spherical. *bid* is a unique body indentifier. *name*, *volume*, *mass* represent the name, mass and volume of the moon respectivly. *escape* represents the speed needed to escape the gravitational field of the moon. *gravity* represents the force of gravity on the moon's surface.

IsMoon(<u>moon</u>, planet)
A tuple in this relation represents the moon and parent planet relationship. *moon* is the moon indentification number of a moon in the orbit of the planet with planet identification number *planet*.

Astronomer(<u>aid</u>, name, nationality)
A tuple in this relation represents an astronomer. *aid* is an unique astronomer identification number. *name* is the astronomer's given name and *nationality* is the nation he identifies with.

Discovered(<u>bid</u>, aid, discovery_date)
A tuple in this relation represents the relationship between a celestial body with identification number *bid* and the individual who discovered that body with identification number *aid*. *discovery_date* is the date in which the body was discovered.

Integrity Constraints

IsMoon[moon] $\subseteq$ (SmallMoon[mid] $\cup$ LargeMoon[mid])
isMoon[planet] $\subseteq$ Planet[pid]
Discovered[astronomer] $\subseteq$ Astronomer[aid]
Discovered[body] $\subseteq$ Planet[pid] $\cup$ SmallMoon[mid] $\cup$ LargeMoon[mid]
π (type) Planet $\subseteq$ {"p", "d"}

Data Dictionary

Table 1: Planet

attribute	discription	type	required	default
bid	the unique id of a planet	int	yes	
name	the name of the planet	string	yes	
type	whether the planet is a dawrf or full size	char	yes	p
mass	the mass of the planet	float	yes	
volume	the volume of the planet	float	yes	
moon_num	the number of moons in the planet's orbit	int	yes	0
escape	the speed needed for a rocket to escape the planet's gravitational field	float	yes	
gravity	the force of gravity on the planet's surface	float	yes	0

Table 2: SmallMoon

attribute	discription	type	required	default
bid	the unique id of a moon	int	yes	
name	the name of the moon	string	yes	
mass	the mass of the moon	float	yes	

Table 3: LargeMoon

attribute	discription	type	required	default
bid	the unique id of a moon	int	yes	
name	the name of the moon	string	yes	
mass	the mass of the moon	float	yes	
volume	the volume of the moon	float	yes	
escape	the speed needed to escape the moon to escape its orbit	float	yes	
gravity	the force of gravity on the moon's surface	float	yes	

Table 4: IsMoon

attribute	discription	type	required	default
moon	the id of the moon whose in the orbit of the planet with bid in the tuple	int	yes	
planet	the id of the planet with the moon in this tuple in the its orbit	int	yes	

Table 5: Astronomer

attribute	discription	type	required	default
aid	the unique id of an astronomer	int	yes	
name	the name of the astronomer	string	yes	
nationality	the nationality of the astronomer	string	yes	

Table 6: Discovered

attribute	discription	type	required	default
body	the unique id of an body	int	yes	
astronomer	the unique id of an astronomer that discovered body	int	yes	
discovery_date	the year the body was discovered	int	yes	

Justification of Design

The original dataset consists of a table that contains all the information. It uses a boolean attribute to differentiate planets and moons, stores every moon's orbit planet as a string, and every planet's moons as a list. In this table, many attributes don't always have a value: some planets and all moons don't have moons, small moons' volume, mass, gravity and escape velocity are zero. Therefore, we categorized the bodies into Planet, SmallMoon and LargeMoon. We use the attribute type for Planet to differentiate between planet and dwarf planet. The reason why we didn't separate it into two tables is that unlike SmallMoon and LargeMoon, planets and dwarf planets share every attribute. In addition, we created the relational key IsMoon to describe the moon and the planet it orbits. This frees "planet" string and "moons" list for SmallMoon, LargeMoon and Planet. Furthermore, instead using "discovered by" and "discovered year" attribute for all bodies, the key Astronomer was defined to store more information of the astronomers, and the relational key Discovered was defined to document all the discoveries. This documents the astronomers and the discoveries in a better way, allows us to explore further into the history of astronomy, and provides insight for our question 1.