Final Project Write-Up

Daniel Justice

3/29/2020

Solar system small body analysis

The Presentation

I have used Python and Jupyter notebooks throughout the course since those are tools that I use daily in my own work. I made a risky decision to change course at "the 11th hour". Jupyter notesbooks has some frustrating limitations when it comes to high quality presentations such as a lack of built-in code hiding and variable templating within Markdown sections. The Python package Pandas draws much of its inspiration from R dataframes in the first place, so it was a natural transition. I am very pleased with the results, and will definitely be using R-Studio going forward for business presentations. The website r-tutor.com authored by Chi Yau was a tremendous help in my learning curve. I have turned off cell echoing to preserve the calculations used in the presentation.

My data set

The data presented here is a sample collected from NASA's JPL Small-Body Database. They represent smaller objects in the Solar System such as comets, asteroids, and moons. I chose this data set because I am curious about space science, and this is an opportunity to learn more about the subject while fulfilling the requirements of the assignment. Please forgive any inaccurate statements about orbital mechanics! There are **22,458 cases** in my sample.

Variables

I learned what most of these variables mean from Holli Riebeek's article on NASA's Earth Observatory Website (Riebeek, 2009).

Title	Description	Type
full_name	Full Name	string
a	[au] semi-major axis	float64
e	eccentricity	float64
i	[deg] inclination	float64
w	[deg] perihelion degree	float64
om	[deg] node degree (longitude of the ascending node)	float64
ma	[deg] M mean anomaly degree	float64
q	[au] perihelion distance	float64
ad	[au] Q aphelion distance	float64
per_y	period years	float64
H	absolute magnitude	float64
moid	[au] Earth MOID (minimum orbit intersection distance)	float64
orbit_id	Orbit solution ID	string
class	Orbit class	string

All of the `float64` data types are **quantitative** variables, and the `string` data types `orbit_id` and `class` are **categorical**.

Orbit Classes

Below is a table of orbit classes. There is a wealth of information on the CNEOS website that provides much more detail than I can here.

class	name	num
AMO	Amor	8267
APO	Apollo	12308
ATE	Aten	1684
ETc	Encke-type Comet	1
HTC	Halley-type Comet	30
IEO	Atira	21
JFC	Jupiter-family Comet*	10
JFc	Jupiter-family Comet	137

JFC -> Jupiter-family comet, classical definition (P < 20 y).
JFc -> Jupiter-family comet, as defined by Levison and Duncan (2 < TJupiter < 3).

```r
mydata = read.csv("../neo.csv")
summary(mydata)
```

Data summary

```
##             full_name           a                  e
##         (1979 XB):    1   Min.    :  0.5554   Min.   :0.002846
##         (1982 YA):    1   1st Qu.:  1.3160   1st Qu.:0.315925
##         (1983 LC):    1   Median :  1.7232   Median :0.459597
##         (1986 NA):    1   Mean   :  1.8372   Mean   :0.446875
##         (1988 NE):    1   3rd Qu.:  2.2082   3rd Qu.:0.572447
##         (1989 AZ):    1   Max.   :453.0215   Max.   :0.997982
##  (Other)         :22452
##        i                   w                   om                 ma
##  Min.   :  0.01352   Min.   :  0.0079   Min.   :  0.026   Min.   :-15.57
##  1st Qu.:  4.62325   1st Qu.: 92.8879   1st Qu.: 82.717   1st Qu.: 54.08
##  Median :  8.95086   Median :184.5539   Median :172.094   Median :164.37
##  Mean   : 12.67136   Mean   :182.1138   Mean   :173.066   Mean   :172.14
##  3rd Qu.: 17.85681   3rd Qu.:271.5291   3rd Qu.:255.628   3rd Qu.:289.22
##  Max.   :172.51374   Max.   :359.9820   Max.   :359.978   Max.   :360.00
##
##        q                  ad                per_y               H
##  Min.   :0.02832   Min.   :  0.6538   Min.   :   0.414   Min.   : 9.40
##  1st Qu.:0.78427   1st Qu.:  1.7179   1st Qu.:   1.510   1st Qu.:20.50
##  Median :0.96387   Median :  2.5009   Median :   2.262   Median :23.00
##  Mean   :0.91446   Mean   :  2.7599   Mean   :   3.323   Mean   :22.82
##  3rd Qu.:1.06960   3rd Qu.:  3.4212   3rd Qu.:   3.281   3rd Qu.:25.10
##  Max.   :1.29999   Max.   :905.1288   Max.   :9642.430   Max.   :33.20
##                                                          NA's   :189
##       moid              orbit_id          class
##  Min.   :0.00000   JPL 3  :  797   APO    :12308
##  1st Qu.:0.01601   JPL 6  :  726   AMO    : 8267
```

```
##  Median :0.05349   JPL 4   :   722   ATE     : 1684
##  Mean   :0.09443   JPL 1   :   715   JFc     :  137
##  3rd Qu.:0.14457   8       :   698   HTC     :   30
##  Max.   :0.70772   9       :   693   IEO     :   21
##  NA's   :14        (Other):18107   (Other):   11
```

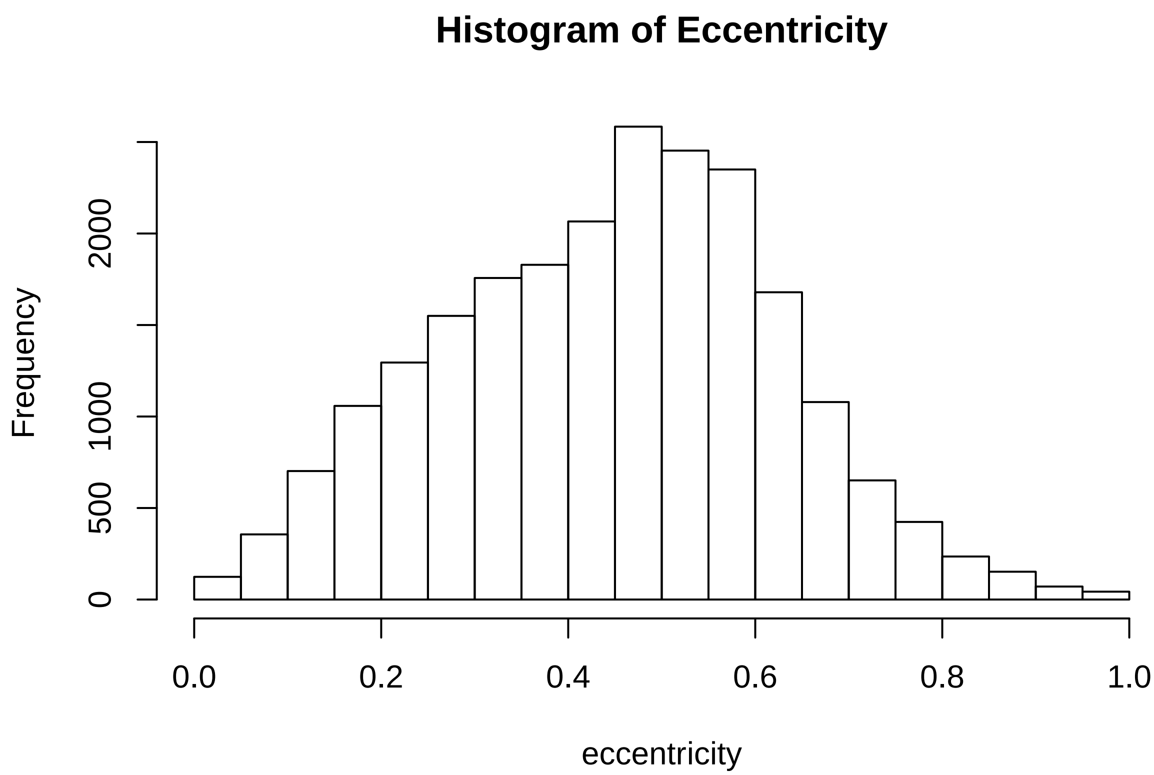
Eccentricity

The first quantitative variable I would like to explore is labeled `e` for eccentricity. This variable takes on values between 0 and 1 where 0 is a perfectly circular orbit and close to 1 flattens the orbit into line (Air Command and Staff College Space Research Electives Seminars, 2009). The distribution appears to be close to **normal** and slightly **skewed to the right**.

```
e_mean = round(mean(mydata$e), digits = 3)
e_median = round(median(mydata$e), digits = 3)
```

The **mean** is 0.447.
The **median** is 0.46.

```
hist(mydata$e, main="Histogram of Eccentricity", xlab = "eccentricity")
```



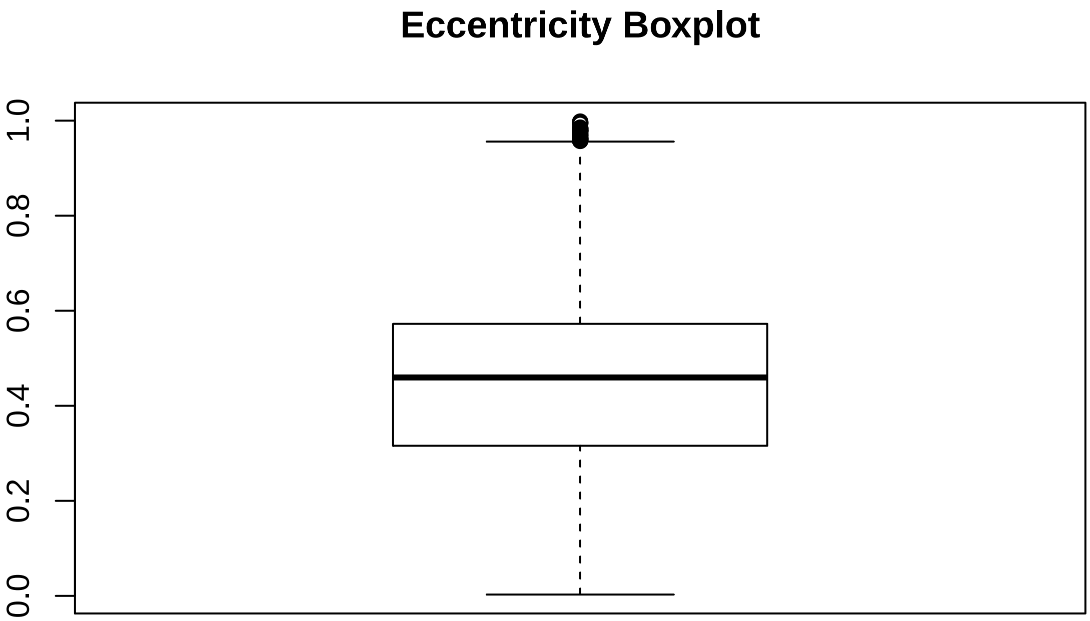

```
summary(mydata$e)
```

```
# Outliers Q3 + 1.5 * IQR > value < Q1 - 1.5 * IQR
q1 = quantile(mydata$e, 0.25)
q3 = quantile(mydata$e, 0.75)
iqr = q3 - q1
low_outliers <- mydata[which(mydata$e < q1 - (1.5 * iqr)),]
high_outliers <- mydata[which(mydata$e > q3 + (1.5 * iqr)),]
```

The **IQR** is 0.257.
The **standard deviation** s is 3.943.

There are 0 **outlier** values less than -0.069
There are 35 **outlier** values greater than 0.957

```
boxplot(mydata$e, main="Eccentricity Boxplot")
```



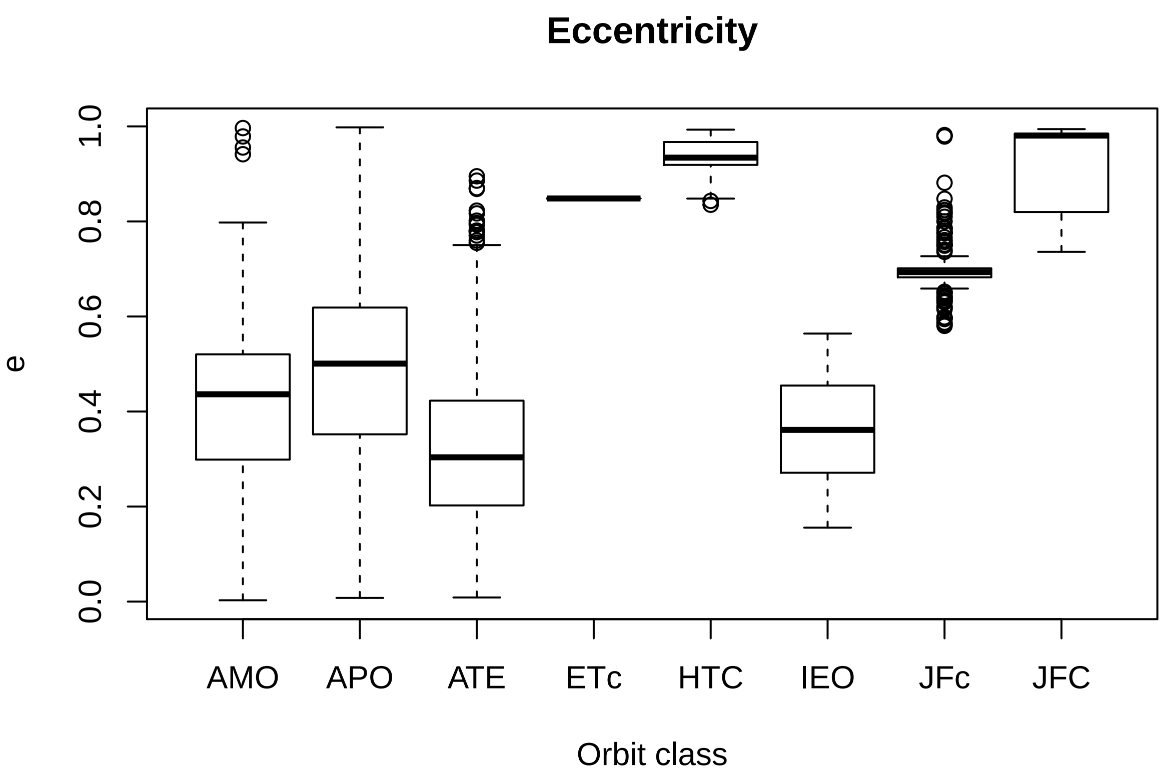
Five-number summary

Variable	Value
minimum	0.003
Q_1	0.316
median	0.46
Q_3	0.572
maximum	0.998

Eccentricity by orbit class

According to NASA, Near-Earth objects (NEO) have orbits that put them in proximity to Earth's orbit. It is not much surprise to see similar IQR's in these classes (AMO, APO, ATE, and IEO). The Atiras class (IEO) is noteworthy for its lack of extreme values. NEO's in this class have orbits contained entirely within Earth's own orbit (NASA).

```
boxplot(mydata$e~mydata$class, main="Eccentricity", xlab="Orbit class", ylab="e")
```



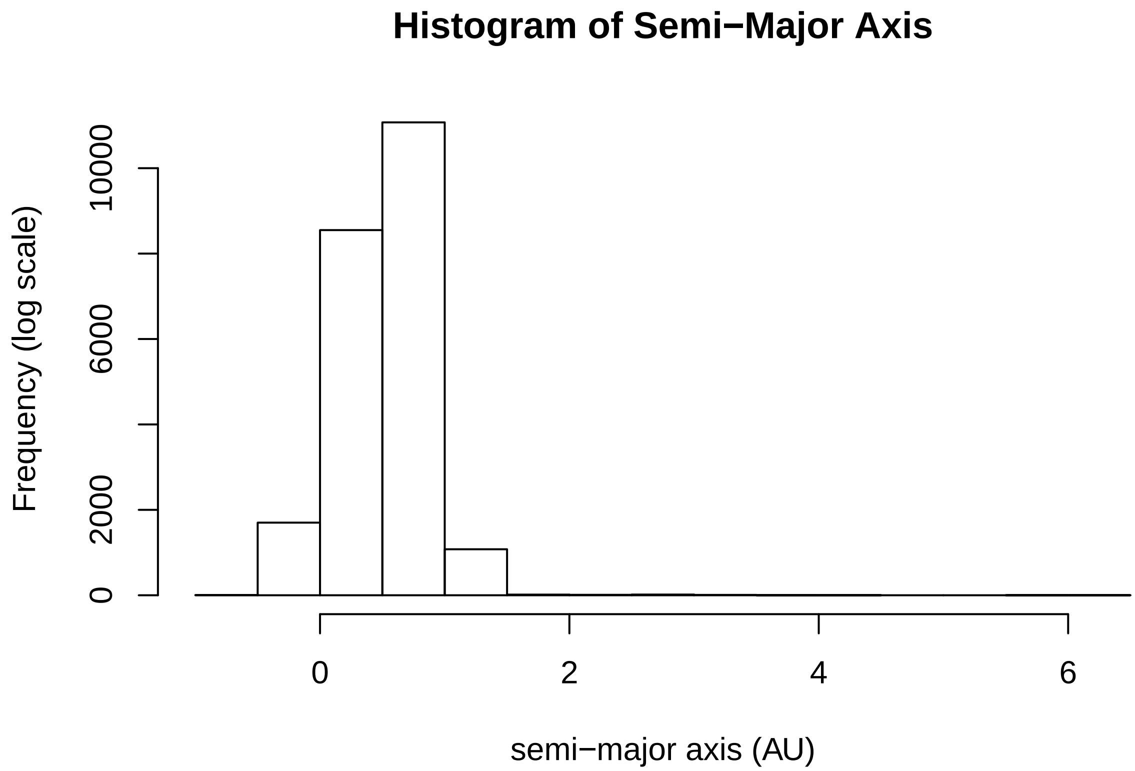
Semi-major axis

The other quantitative variable I would like to explore is the semi-major axis. This is the larger of the two axis of an ellipse (for `e` > 0) (Air Command and Staff College Space Research Electives Seminars, 2009). This distribution has a log-normal shape and is **skewed to the right** with a much longer tail than the previous graph. Given the spread of this data, please note the *logarithmic scale* where indicated.

```
a_mean = round(mean(mydata$a), digits = 3)
a_median = round(median(mydata$a), digits = 3)
```

The **mean** is 1.837.
The **median** is 1.723.

```
hist(log(mydata$a), main="Histogram of Semi-Major Axis", xlab = "semi-major axis (AU)", ylab="Frequency
```



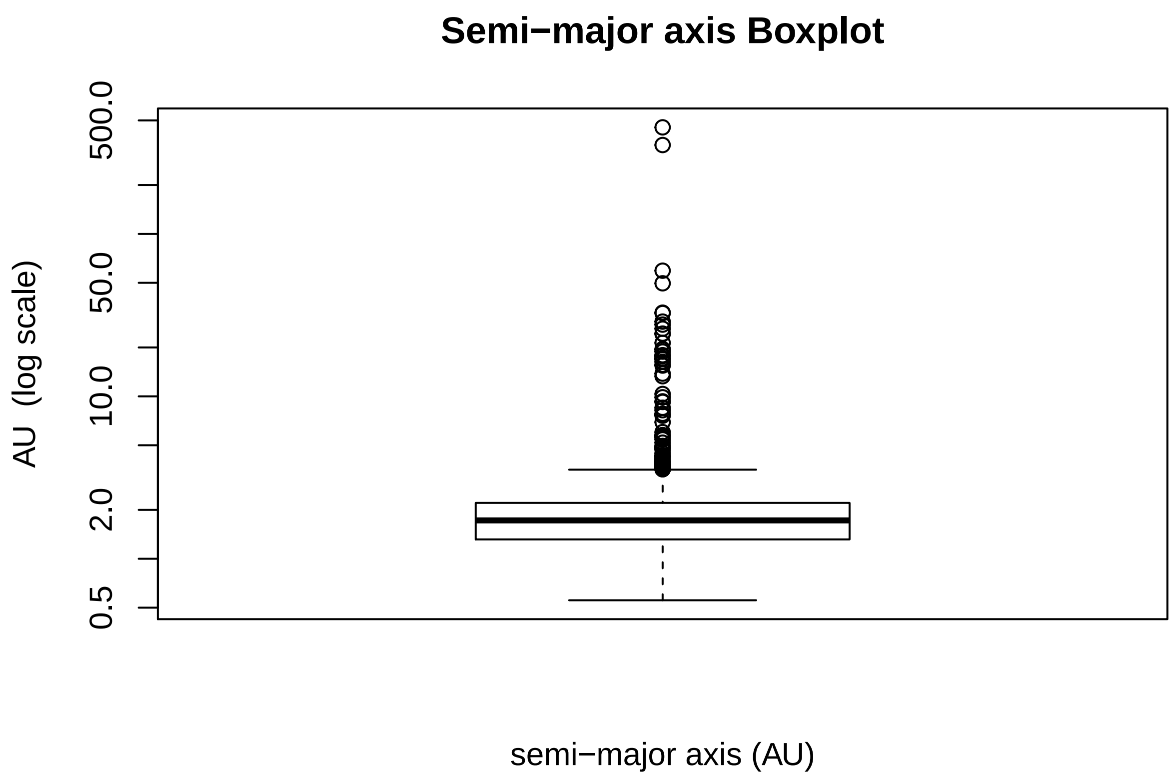

```
summary(mydata$a)
```

```
# Outliers Q3 + 1.5 * IQR > value < Q1 - 1.5 * IQR
q1 = quantile(mydata$a, 0.25)
q3 = quantile(mydata$a, 0.75)
iqr = q3 - q1
low_outliers <- mydata[which(mydata$a < q1 - (1.5 * iqr)),]
high_outliers <- mydata[which(mydata$a > q3 + (1.5 * iqr)),]
```

The **IQR** is 0.892.
The **standard deviation** s is 3.943.

There are 0 **outlier** values less than -0.022
There are 98 **outlier** values greater than 3.546

```
boxplot(mydata$a, main="Semi-major axis Boxplot", log="y", xlab="semi-major axis (AU)", ylab="AU  (log s
```



Five-number summary

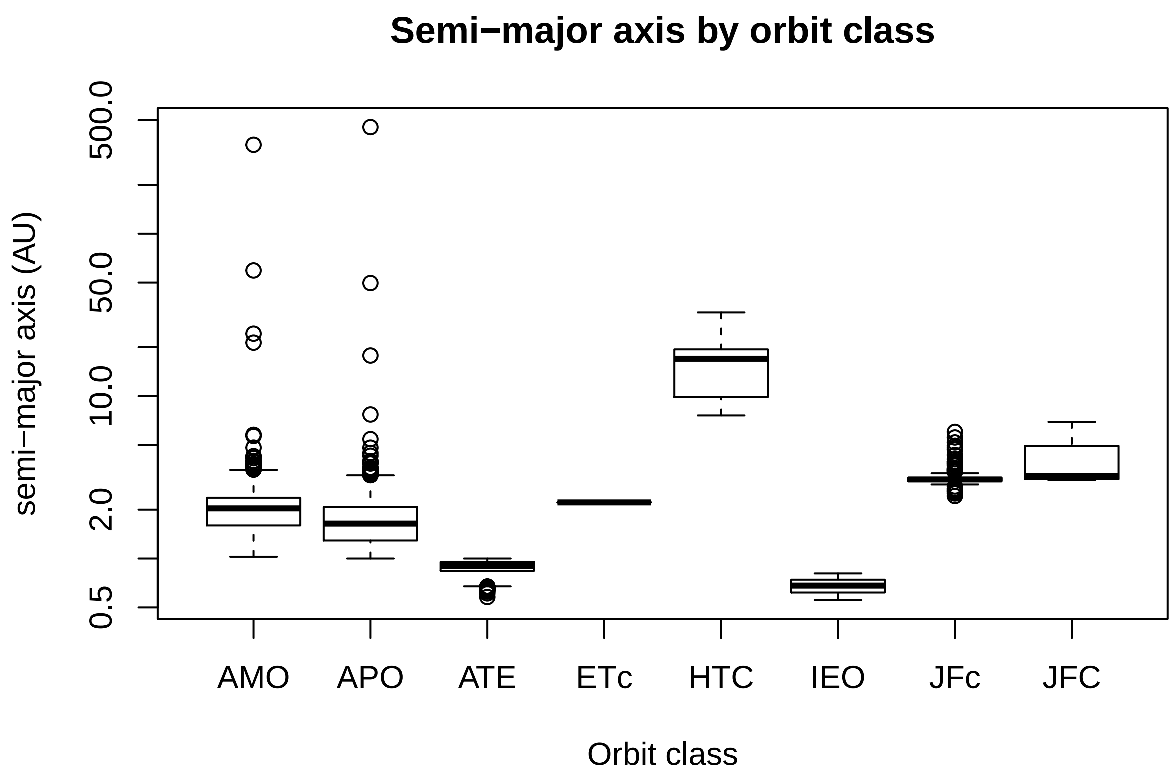
Variable	Value
minimum	0.555
Q_1	1.316
median	1.723
Q_3	2.208
maximum	453.022

Semi-major axis by orbit class

Aten (ATE) and Atira (IEO) class bodies have semi-major axes smaller than 1 AU by definition. I am surprised that there are no outliers in the Jupiter and Haley-family classes. This may be due to observational practices or other constraints that I am not familiar with.

For reference, Earth is 1 AU from the sun, and Pluto is around 40 AU from the sun. Some of these bodies travel very far outside of our planetary system!

```
boxplot(mydata$a~mydata$class, main="Semi-major axis by orbit class", xlab="Orbit class", ylab="semi-ma
```

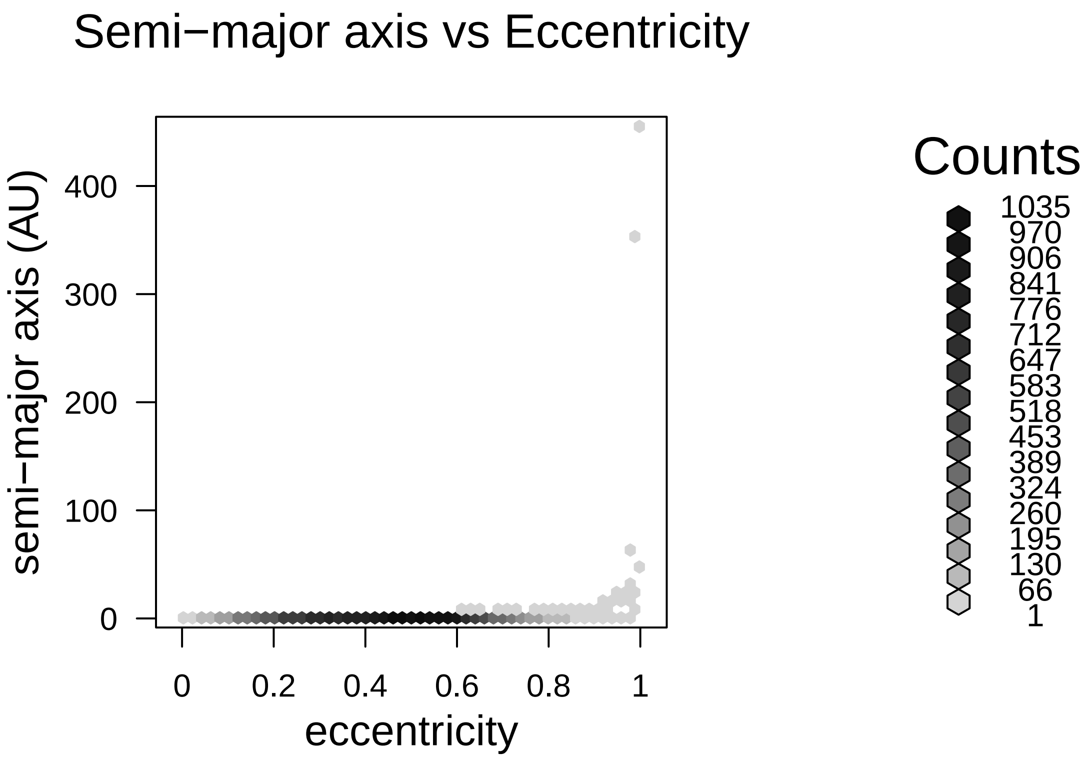


Linear correlation

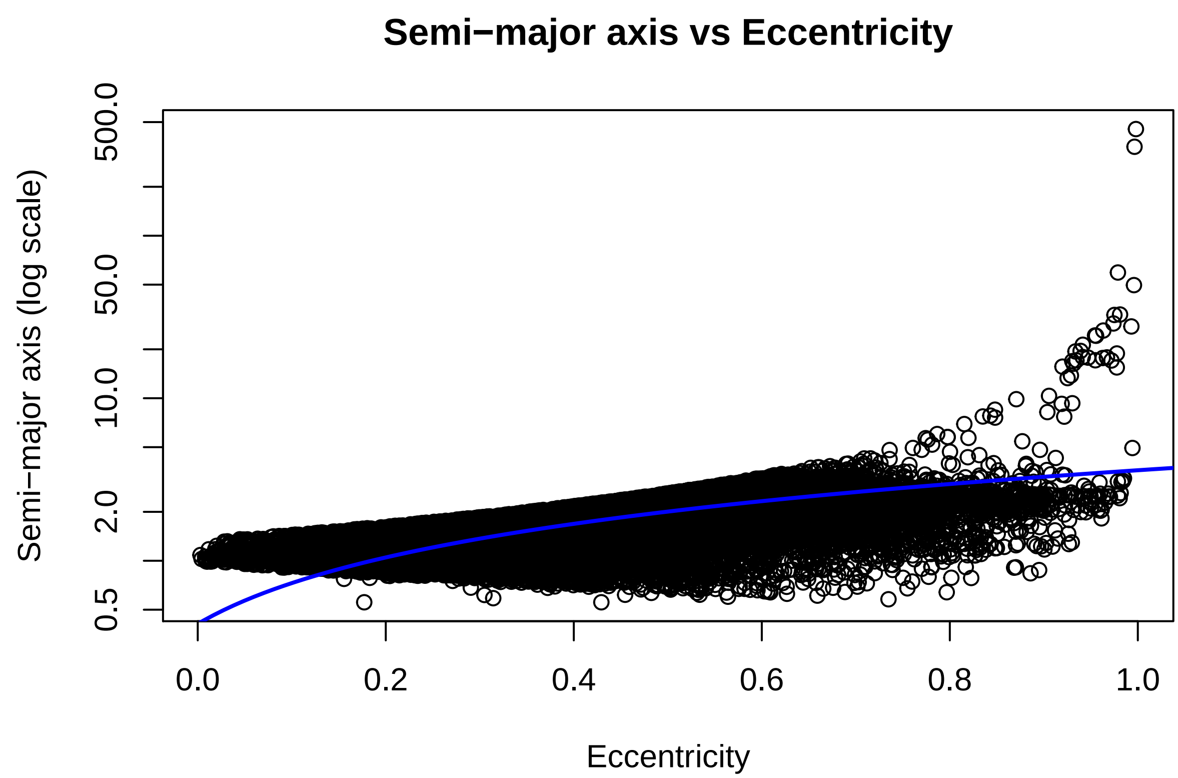
Scatter plots can lose their usefulness with too many data points. I read about a plot called a “hex bin”, and I feel this did a much better job of allowing the data to tell its story (“Quick-R: Scatterplots,” n.d.).

```
library(hexbin)
# plot(mydata$e, mydata$a, log="y", ylab="semi-major axis", xlab="eccentricity")

hbin <- hexbin(mydata$e, mydata$a, xbins=50)
plot(hbin, main="Semi-major axis vs Eccentricity", ylab="semi-major axis (AU)\n", xlab="eccentricity")
```



I struggled with the decision to use a logarithmic scale on a graph depicting a best-fit "line". Ultimately, I feel using this scale reveales much more about the data than a clumb of color along the x-axis.



The **correlation coefficient** r is 0.145. One possible **confounding, or lurking** variable is `i` or inclination. Inclination exhibits a similar correlation coefficient and clustering of values. The r-value is low, but it appears from the clustering in the above graph that one would have a decent probability of finding satellites near the regression line, especially for larger values of eccentricity. I am surprised that orbits with large semi-major axes tend to be highly elliptic as well.

Summary

I do not have the physics background and experience to do this data justice. There are many more variables and relationships that could be explored much further. I am finishing this assignment with many more questions than answers about this data set, but it was an exciting opporunity to explore a new subject! I incurred a bit of extra work by making the switch to R, but I am very happy with the results, and I will be able to use this experience on the job in the near future.

References

- Air Command and Staff College Space Research Electives Seminars. (2009, September 1). AU-18 Space Primer. Retrieved March 29, 2020, from https://www.airuniversity.af.edu/Portals/10/AUPress/Books/AU-18.PDF
- NEO Basics. (n.d.). Retrieved March 29, 2020, from https://cneos.jpl.nasa.gov/about/neo_groups.html
- Quick-R: Scatterplots. (n.d.). Retrieved March 29, 2020, from https://www.statmethods.net/graphs/scatterplot.html
- Riebeek, H. (2009, September 4). Catalog of Earth Satellite Orbits. Retrieved March 29, 2020, from https://earthobservatory.nasa.gov/features/OrbitsCatalog

Additional sites used:

- https://ssd.jpl.nasa.gov/sbdb_query.cgi
- https://cneos.jpl.nasa.gov/glossary/