Statistical Analysis of Near Earth Objects

Danielle Bisordi Kristin Buckley

December 20, 2009

Abstract

In this paper, we explore the statistical properties of Near Earth Objects (NEOs). It is determined whether the date has an effect on the quantity of NEOs observed, as well as if the velocities, sizes and orbit types of the NEO are correlated with the miss distances.

1 Introduction

Near Earth Objects (NEOs) are comets and asteroids that have been nudged by the gravitational attraction of nearby planets into orbits that allow them to enter in the Earth's neighborhood. They are composed mainly of water and ice and are remnants of the formation of the solar system about 4.6 billion years ago.[1] NEOs pose a threat to Earth since there is a possibility an NEO could one day strike Earth. This would disrupt life on Earth in unpredictable ways, possibly destroying all life on Earth if the NEO is large enough; therefore statistically analyzing their properties is very useful, both for astrophysical theory about the formation and evolution of solar systems, and also for predicting patterns in NEOs which could help predict potentially hazardous impacts. The data used in this analysis consists of the 11,726 NEOs observed by NASA in the past 10 years.

The data types used in this study are year, month, miss distance, velocity, absolute magnitude, and class. Year and month refer to the year and month in which the NEO was closest to Earth. The miss distance is the smallest value of the distance between the Earth and the NEO. The absolute magnitude (H)is the value astronomers measure to indirectly evaluate the diameter of the NEO. Since the diameter cannot be directly measured (because it is in space), indirect measures must be taken. In this case, the relationship between the absolute magnitude and diameter is inversely proportional; the lower the H value, the larger the size of the NEO.[2] Finally, the class of the NEO describes the orbit of the NEO. Atira (IEO) is an "Inner Earth Orbit" NEO, whose orbit is contained completely within the orbit of the Earth, that is, the NEO is always found between the Earth's orbit and the Sun. Atens (ATE) NEOs have very elliptical orbits that pass through Earth's orbit, and whose semi-major axes is smaller than Earth's. Apollo (APO) NEOs also have orbits that pass through

Earth's orbit, but have a semi-major axis that is larger than Earth's. Amor (AMO) NEOs have orbits that are completely contained between Earth and Mars.[3] To keep the data more consistent, we chose our data to have the group with the smallest H values (largest diameters) that the NASA site offered. This ensures that over the ten years of observation, the NEOs we chose were more or less equally observable, and would be detected in the early years of the this period as well as the later years, regardless of technological advances.

2 Statistical Analysis

2.1 Does the year have an effect on the quantity of near misses?

Figure 1 contains a sample of the data we are working with. For each close approach of a near earth object the data provides us with a month and year in which this close approach occurs. The first thing we wanted to understand from this data is whether these close approaches occurred regularly throughout each year or if they were increasing in occurrence each year. To determine which of our hypotheses were correct we created a histogram to help us visually compare the relationship between the year and the frequency in occurrence of near earth objects each year.

The histogram in Figure 2 visually confirms our hypothesis that the frequency of near earth object approaches is increasing each year in what appears to be a linear fashion. Our next step was to confirm this apparent linear increase in NEO approaches. While our histogram appears to show a linear relationship to be sure our eyes were not deceiving us we performed a linear regression test using the year as our response variable. We compared that with the frequency of NEO occurrence each year. From this linear regression test we found that 95.87 percent of the NEO approach frequency data per year can be described by a linear relationship. See Appendix A for the full test results. The graph in Figure 3 plots the frequency of NEO occurrences per year and the red line is the linear regression model of the data.

The plot in Figure 3 shows more clearly the linear relationship between year and frequency of occurrence of NEOs.

2.2 Does the month have an effect on the quantity of near misses?

Since our data showed such a strong linear relationship between year and frequency of NEO occurrence we wanted to explore more deeply the occurrence of NEOs within a year's time to see if we could find any other interesting patterns. Again we consulted a histogram to see if any obvious patterns emerged from the data we had. The Figure 4 shows the result of the frequency of NEO appearances and the month in which they appeared.

This histogram presented to us another very interesting pattern in our data. It seems as though more NEO appearances happen in January than any other month in the year. The next month in which the most NEO appearances occur is October, however there are 1000 more NEO appearances in January than there are in October.

2.3 Do the absolute magnitude and the relative velocity of the NEO have an effect on the miss distance?

The next thing we wanted to learn about our data is whether or not the absolute magnitude and the velocity played a statistically significant part in determining the miss distance of the NEO. The first step we made to determine this was to plot the data and look for any type of relationship the data might have. Figures 5 and 6 show the graphs of the miss distance vs. the velocity of the NEO and the miss distance vs. the absolute magnitude of the NEO.

As we can see from the plots, the velocity of the NEOs seems to have a linear relationship with the miss distance while the absolute magnitude of the NEOs seems to have more of a logarithmic relationship with the miss distance.

To determine whether each of these factors have a statistically significant effect on the miss distance we performed three separate tests. We first used a linear regression test to determine whether the velocity of an NEO is a significant factor in determining the miss distance. From this test we were able to see that the velocity of a NEO does in fact have an effect on the miss distance of that NEO within a margin of error of less than .001 percent. Second, we use a logistic regression test to determine if the absolute magnitude of a NEO is a significant factor in determining the miss distance of that NEO. This test revealed that the absolute magnitude of a NEO also plays a statistically significant roll in determining the miss distance of an NEO within a margin of error of less than .001 percent. Finally, we performed a multivariate linear regression test to determine whether both the velocity and absolute magnitude of a NEO play a separate role in determining the miss distance of that NEO or if they act individually in effecting the miss distance of the NEO. From this test we again found that within less than a .001 percent margin of error the velocity and absolute magnitude of a NEO are statistically significant in determining the miss distance of a NEO. See appendix B for full test results.

2.4 Does the class of NEO have an effect on the miss distance, velocity and absolute magnitude of the NEO?

Now that we have determined that both the velocity and absolute magnitude have an effect on the miss distance we are curious to see how these factors compare to each other from class to class of NEO. Since the class of NEO is determined by the type of orbit the NEO travels in our analysis will show whether each type of orbit contains a variety of NEOs with different velocities, sizes, and miss distances or if higher velocities, sizes and miss distances are all in one orbit while the NEOs with lower values of these traits are all in another orbit.

2.4.1 Does the class of NEO have an effect on the miss distance?

To determine whether the class of NEO, and thus the orbit of the NEO, has an effect on the miss distance we first visually examine the average miss distance per class to see if we can determine a noticeably different mean miss distance per class of NEO. To do this we created a box plot that enables us to view the mean miss distance and spread of the miss distances for each class side by side.

From the box plot in Figure 7 we can plainly see that the IEO class of NEOs has an overall larger average miss distance than any other class of NEO. We can also see that the APO class has a much smaller average miss distance than any other class of NEO. We then performed an ANOVA test and found that the class of NEO does in fact have a statistically significant effect on the NEO miss distance by a margin of error of less than .001 percent. See appendix C for full test results.

2.4.2 Does the class of NEO have an effect on the velocity?

To determine the effect of NEO class on the velocity of the NEO we follow the same procedure. First we created a box plot to visually analyze the effects of NEO class on the velocity of the NEO.

The box plot in Figure 8 shows that the IEO class of NEOs has the highest average velocity while the AMO class has the lowest average velocity. Next we performed an ANOVA test and from this test we found that within .001 percent margin of error the class of an NEO is statistically significant in effecting the velocity of that NEO. See appendix C for full test results.

2.4.3 Does the class of NEO have an effect on the absolute magnitude?

Finally, to determine if the class of NEO has any effect on the absolute magnitude of that NEO we again performed the same tests as we did for velocity and miss distance. The box plot in Figure 9 shows that the IEO class of NEOs has the lowest average absolute magnitude and the ATE class has the highest average absolute magnitude.

Next we performed an ANOVA test to determine whether the class of NEO has an effect on the absolute magnitude of the NEO. And again we found that the class of an NEO is statistically significant in effecting the absolute magnitude of an NEO within less than a .001 percent margin of error. See appendix C for full test results.

2.5 Assuming there are more NEOs in existence than our sample data describes, is it likely that there are any NEOs that will have a miss distance of less than .0026 AU?

Now that we have determined that all of the factors that we are dealing with do interact with each other significantly and we know a little bit about how they affect each other we can use this information to make an accurate prediction about the future NEOs. The ultimate question we wanted to answer in our analysis was whether one day one of these NEOs will hit earth or at least come close enough to significantly effect Earth. We know that the moon has a significant effect on Earth and that it is .0026 AU away. Now we want to see if it is statistically possible for an NEO to travel that close to Earth in the future assuming that the past NEO trends continue in the future. In order to determine this we performed an upper tailed t test to determine whether or not the mean miss distance was statistically significantly different from or less than .0026 AU. We expected this test to show that we could reject our null hypothesis and that the average miss distance is in fact greater than .0026AU and the t test confirmed our suspicions. See appendix D for full test results. Next we calculated an 99 percent confidence interval to see if it there was a possibility that an NEO could travel that close to earth in the future. From this calculation we can say with 99 percent confidence that an NEO will never travel on average closer to the Earth than .2308 AU.

3 Conclusions

From our analysis it can be seen that the number of near misses seems to be increasing with each year. While this is troubling, this may be attributed to better technology detecting more and more NEOs each year. This was taken into account in our data selection, as noted in the introduction, but because there is still a wide range of sizes in our data (which there must be for the analysis involving size to be accurate), technology still could play a factor in this.

The observation that there are many more NEOs in January than in any other month can be attributed to the location of Earth in its orbit during the month of January. This may have to do with the fact that Earth reaches its perihelion during January each year, which is the part of its orbit where it is closest to the sun. Since many NEOs travel in highly elliptical orbits and orbit extremely close to the sun when they approach it, this may explain the sudden increase in NEOs during the month of January.

The velocity of the NEOs had a linear relationship with the miss distance. This may also be explained by the proximity to the sun. The gravitational pull of the sun may be causing the NEOs to travel away from Earth faster the closer they are to the sun (and therefore, farther away from Earth) they are. This can be seen in Figure 5, where it is obvious than many lower velocities are also located far away from Earth; these NEOs may be located on the other side of Earth (farther away from the sun, but equally as far away from Earth as the NEOs closer to the sun).

The size of the NEO had a logarithmic relationship with the miss distance, where the larger the NEO the larger the miss distance. The very small NEOs seemed to be located very close to Earth, and while the larger the NEOs got, the farther away they were at their closest point of approach, they did not get farther away as "quickly" in the plot as the smaller NEOs did as they got closer to Earth. Part of this may be that the closer to Earth the NEO, the easier they are to detect, so a higher quantity of the small NEOs close to Earth are seen than NEOs of the same size farther away from Earth. Another explanation might be that the smaller NEOs more easily get caught in Earth's gravitational field, so approach Earth more closely. The extremely close NEOs are also probably small because they start to get broken up by the Earth's atmosphere.

The IEO NEOs had the largest miss distances and the APO NEOs had the smallest miss distance. The APOs are only found between the orbits of Mars and Earth. This information can be used in determining which NEOs to track more carefully when looking for possible threats.

The IEOs also had the highest velocity. Similar to the analysis of the velocities relationship with miss distance, this may be due to the fact that the IEOs are closer to the Sun than any other class of NEO for its entire orbit and this may influence the velocity of the NEOs of this type.

The IEO class also has the lowest absolute magnitude - which means they have the largest average diameter. This makes sense since it was also seen that they have the largest miss distance, and it was determined earlier that the larger the NEO, the larger the miss distance.

Finally, the distance from the Earth to the Moon is approximately .0026AU. If a large NEO were to come this close to Earth, it could effect life on Earth without even impacting Earth, since it could effect the net gravitational pull on Earth, effecting tides and the weather (much like the gravity of the moon does). This is why the distance of .0026AU was chosen to study. If the NEOs continue to behave in the way they have for the last 10 years, it seems that there is very little chance that one will come that close to Earth, which is relevant to risk assessment of the known NEOs.

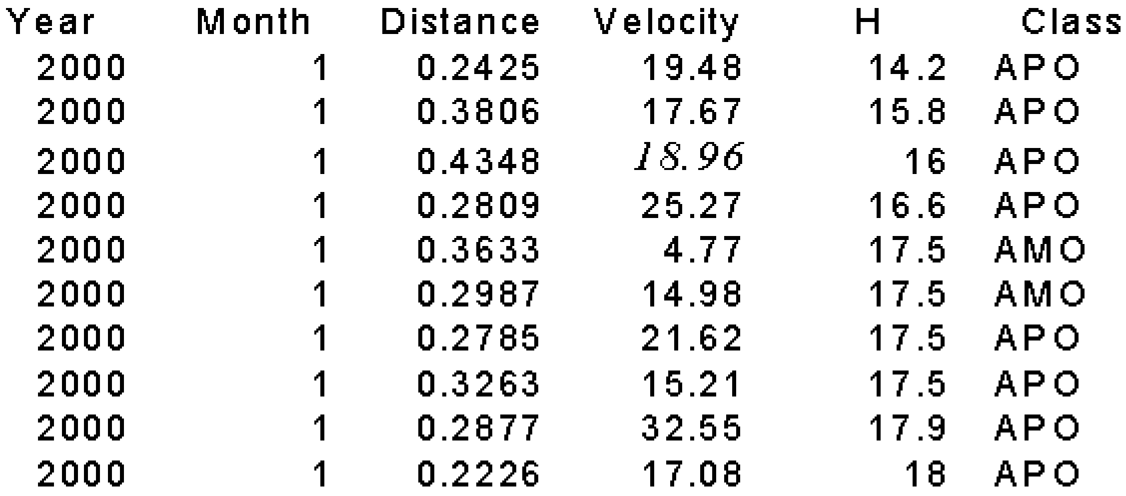
References

[1] http://eo.jpl.gov/neo

[2] http://neo.jpl.nasa.gov/glossary/h.html

[3] http://neo.jpl.nasa.gov/neo/groups.html

Figures:



Year	Month	Distance	Velocity	H	Class
2000	1	0.2425	19.48	14.2	APO
2000	1	0.3806	17.67	15.8	APO
2000	1	0.4348	*18.96*	16	APO
2000	1	0.2809	25.27	16.6	APO
2000	1	0.3633	4.77	17.5	AMO
2000	1	0.2987	14.98	17.5	AMO
2000	1	0.2785	21.62	17.5	APO
2000	1	0.3263	15.21	17.5	APO
2000	1	0.2877	32.55	17.9	APO
2000	1	0.2226	17.08	18	APO

Figure 1: A sample of the data

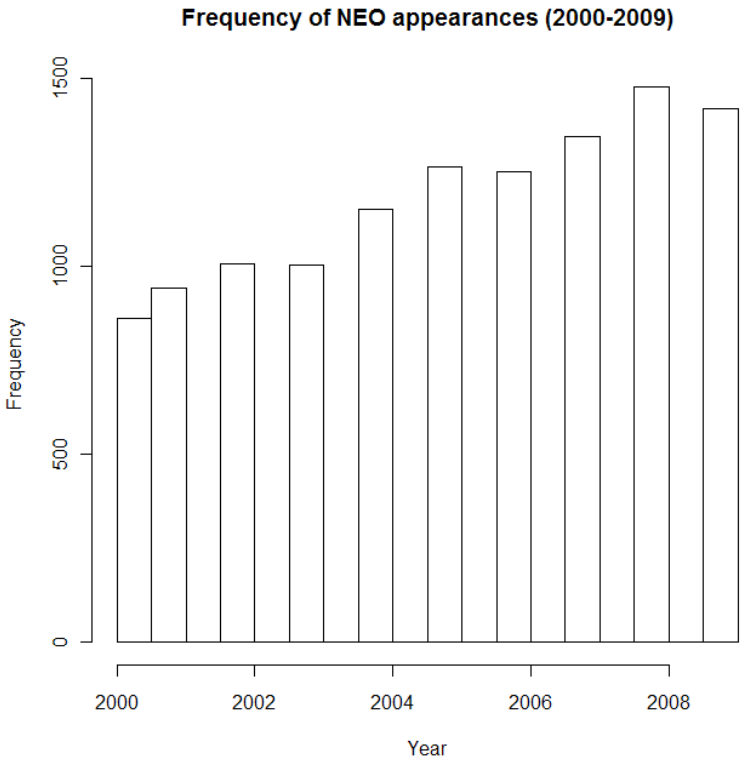


Figure 2: Histogram of the frequency of NEOs by year

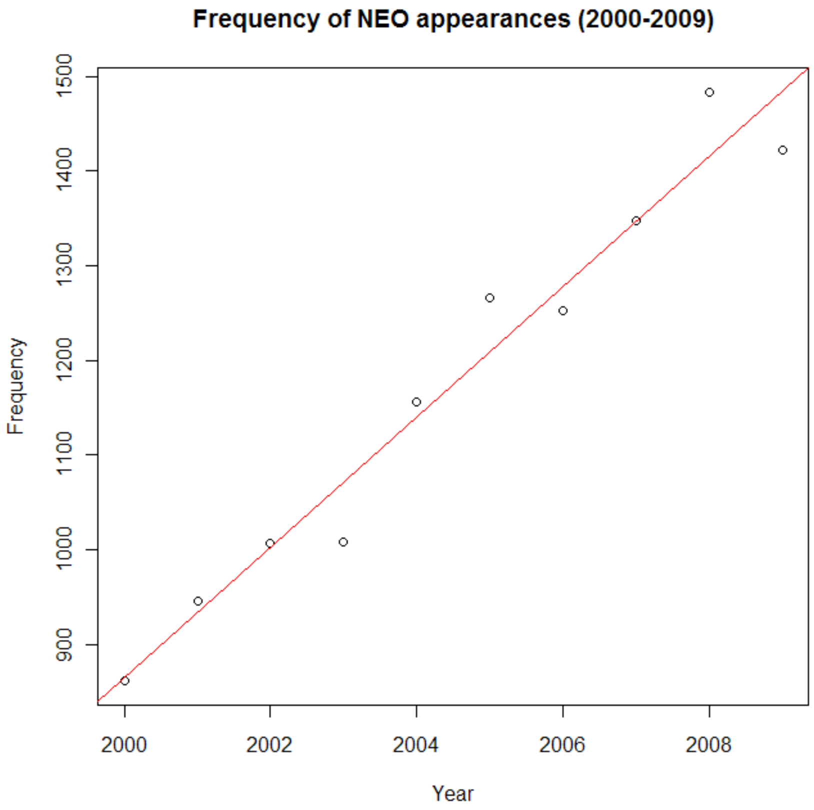


Figure 3: Linear Regression of NEO frequency by year

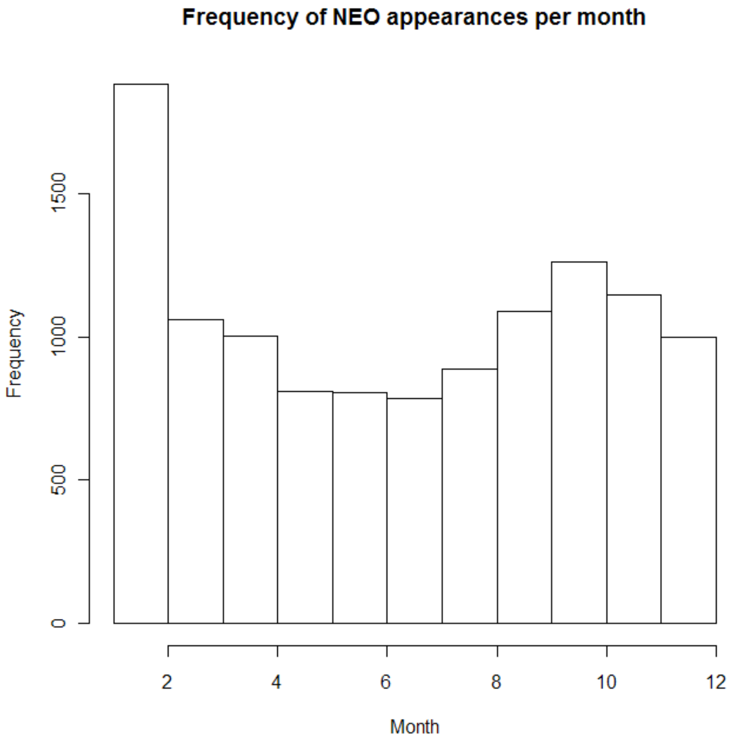


Figure 4: Histogram of the NEO frequency by month

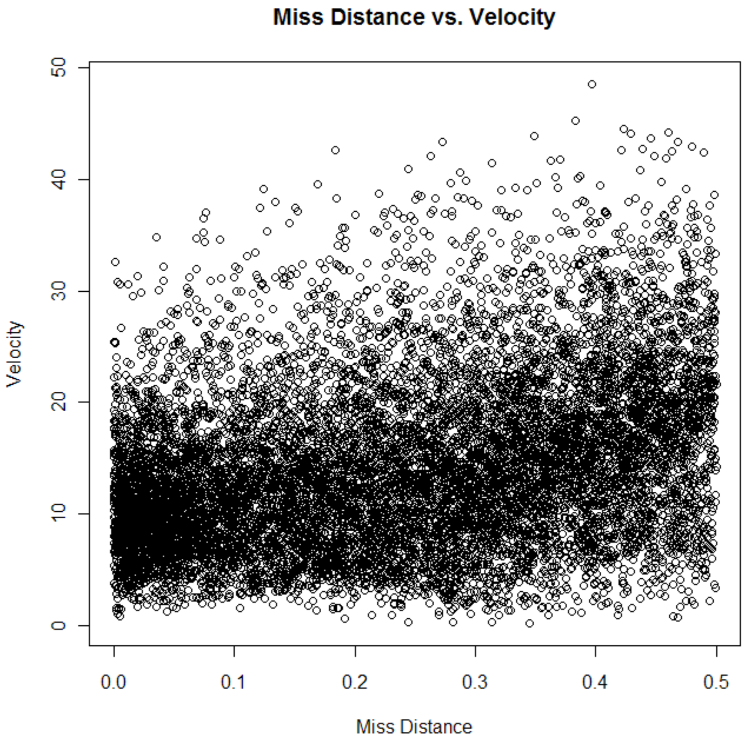


Figure 5

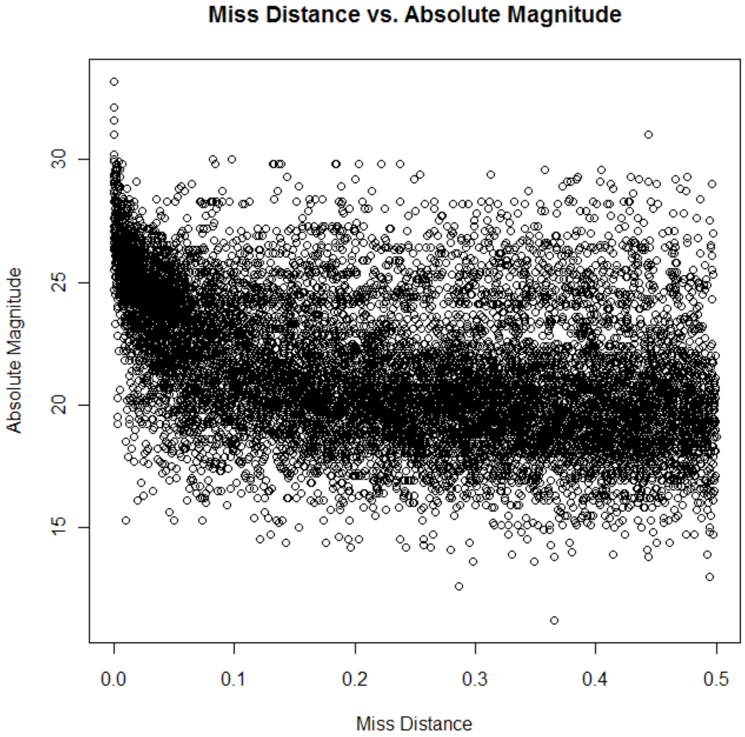


Figure 6

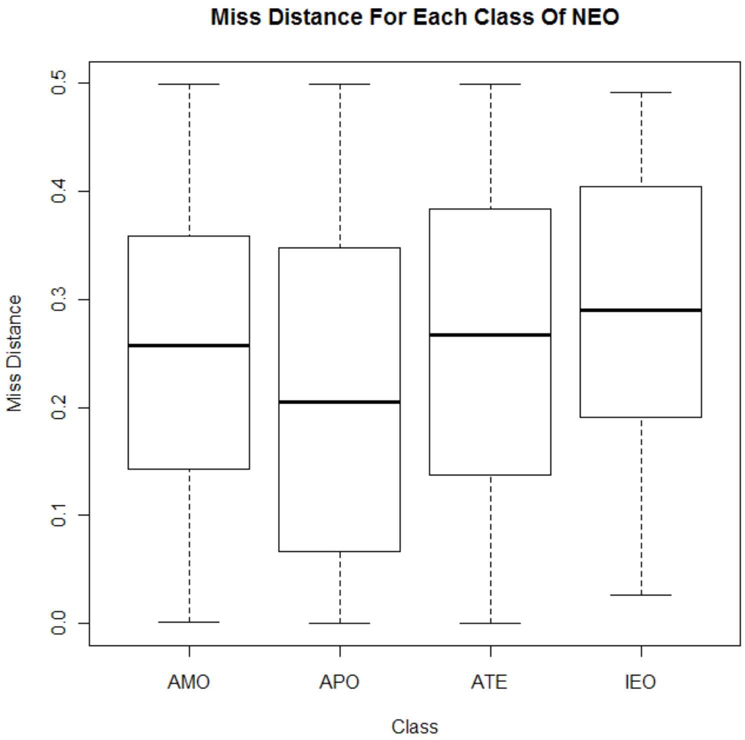


Figure 7: Box plot of miss distances for each class of NEO

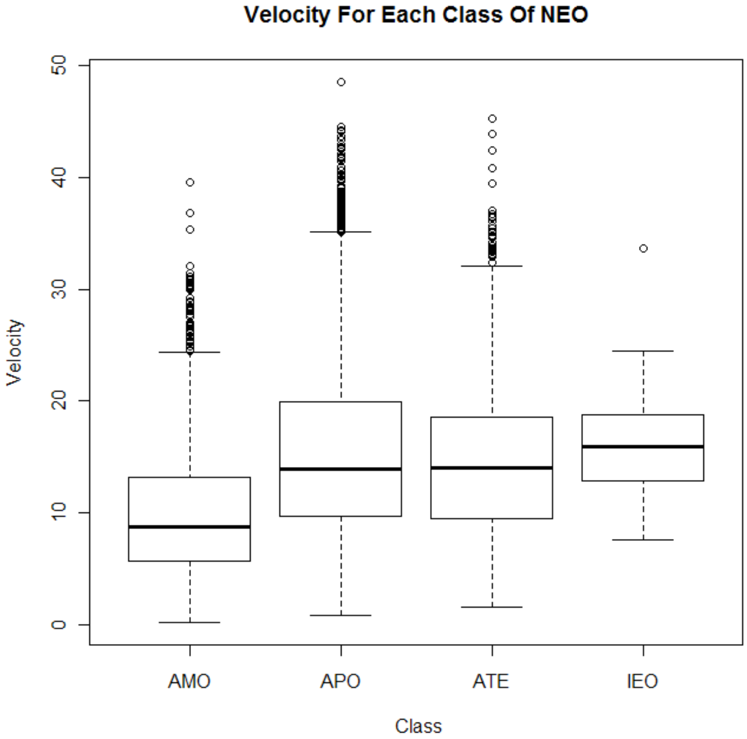


Figure 8: Box plot of velocities for each class of NEOs

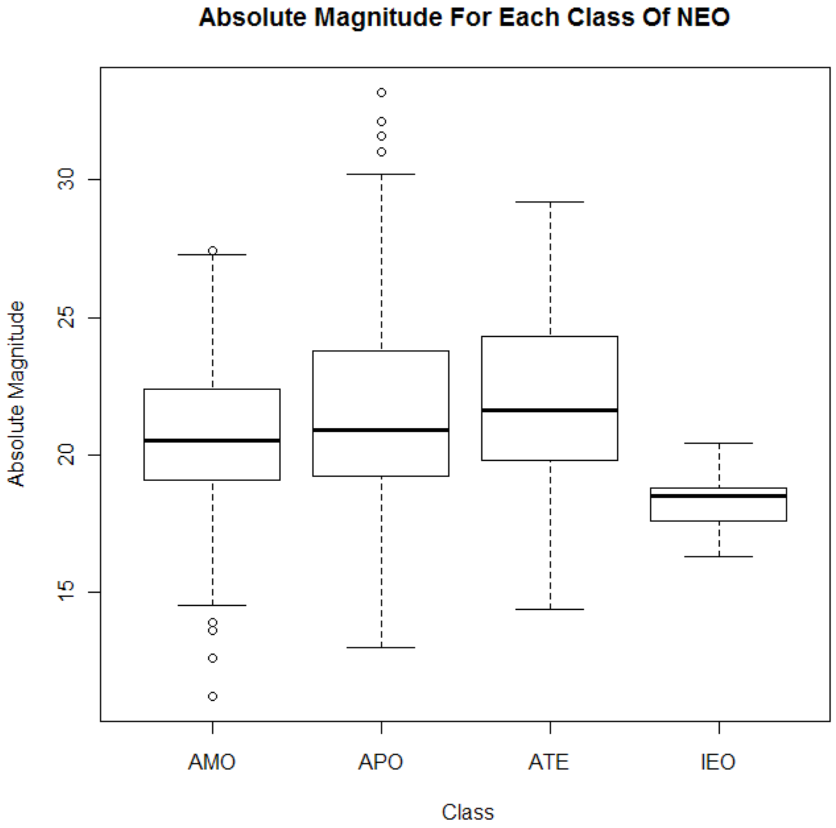


Figure 9: Box plot of absolute magnitudes for each class of NEOs

Appendix A:

Linear regression test of year vs. frequency of occurrence:

```
> summary(lm(2000:2009~yearfreq))

Call:
lm(formula = 2000:2009 ~ yearfreq)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.8015 -0.2847 -0.1476  0.3361  1.0613 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.988e+03  1.217e+00 1633.07  < 2e-16 ***
yearfreq    1.392e-02  1.021e-03   13.63 8.09e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 

Residual standard error: 0.6526 on 8 degrees of freedom
Multiple R-squared: 0.9587,     Adjusted R-squared: 0.9535 
F-statistic: 185.7 on 1 and 8 DF,  p-value: 8.09e-07
```

Appendix B:

Linear regression test of NEO miss distance vs. NEO velocity:

```
> summary(lm(neodata$Distance~neodata$Velocity))

Call:
lm(formula = neodata$Distance ~ neodata$Velocity)

Residuals:
       Min         1Q     Median         3Q        Max
-0.3565282 -0.1238339  0.0008694  0.1154380  0.3391019

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      0.1402413  0.0027850   50.36   <2e-16 ***
neodata$Velocity 0.0066622  0.0001754   37.98   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1395 on 11724 degrees of freedom
Multiple R-squared: 0.1095,     Adjusted R-squared: 0.1095
F-statistic:  1442 on 1 and 11724 DF,  p-value: < 2.2e-16
```

Logistic regression test of NEO miss distance vs. NEO absolute magnitude:

```
> summary(glm(neodata$Distance~neodata$H, family=binomial))

Call:
glm(formula = neodata$Distance ~ neodata$H, family = binomial)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.94339  -0.32659  -0.03367   0.23286   0.99977

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.518324   0.169312   8.968   <2e-16 ***
neodata$H   -0.127808   0.008039 -15.898   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)
```

```
    Null deviance: 1620.1  on 11725  degrees of freedom
Residual deviance: 1352.0  on 11724  degrees of freedom
AIC: 6337.4

Number of Fisher Scoring iterations: 4
```

Multivariate linear regression test between NEO miss distance, velocity, and absolute magnitude:

```
> summary(lm(neodata$Distance~neodata$Velocity+neodata$H))

Call:
lm(formula = neodata$Distance ~ neodata$Velocity + neodata$H)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.37527 -0.11008 -0.01046  0.10265  0.41578 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)       0.5594013  0.0110211   50.76   <2e-16 ***
neodata$Velocity  0.0038801  0.0001796   21.60   <2e-16 ***
neodata$H        -0.0177219  0.0004526  -39.15   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1312 on 11723 degrees of freedom
Multiple R-squared: 0.2125,     Adjusted R-squared: 0.2124 
F-statistic:  1582 on 2 and 11723 DF,  p-value: < 2.2e-16
```

Appendix C:

ANOVA test of NEO miss distance vs. class of NEO:

```
> anova(lm(neodata$Distance~neodata$Class))
Analysis of Variance Table

Response: neodata$Distance
                Df  Sum Sq Mean Sq F value    Pr(>F)
neodata$Class     3   4.816   1.605  74.895 < 2.2e-16 ***
Residuals     11722 251.246   0.021
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

ANOVA test of NEO velocity vs. class of NEO:

```
> anova(lm(neodata$Velocity~neodata$Class))
Analysis of Variance Table

Response: neodata$Velocity
                Df Sum Sq Mean Sq F value    Pr(>F)
neodata$Class     3  56188   18729  381.33 < 2.2e-16 ***
Residuals     11722 575735      49
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

ANOVA test of NEO absolute magnitude vs. class of NEO:

```
> anova(lm(neodata$H~neodata$Class))
Analysis of Variance Table

Response: neodata$H
                Df Sum Sq Mean Sq F value    Pr(>F)
neodata$Class     3   2160     720  86.657 < 2.2e-16 ***
Residuals     11722  97376       8
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Appendix D:

Upper tailed t test of NEO miss distance with the null hypothesis being that the average miss distance is less than or equal to .0026AU:

```
> t.test(neodata$Distance, alternative= "greater", mu=.0026, conf.level=.99)
```

```
	One Sample t-test

data:  neodata$Distance
t = 169.5705, df = 11725, p-value < 2.2e-16
alternative hypothesis: true mean is greater than 0.0026
99 percent confidence interval:
 0.2308399       Inf
sample estimates:
mean of x
0.2340152
```