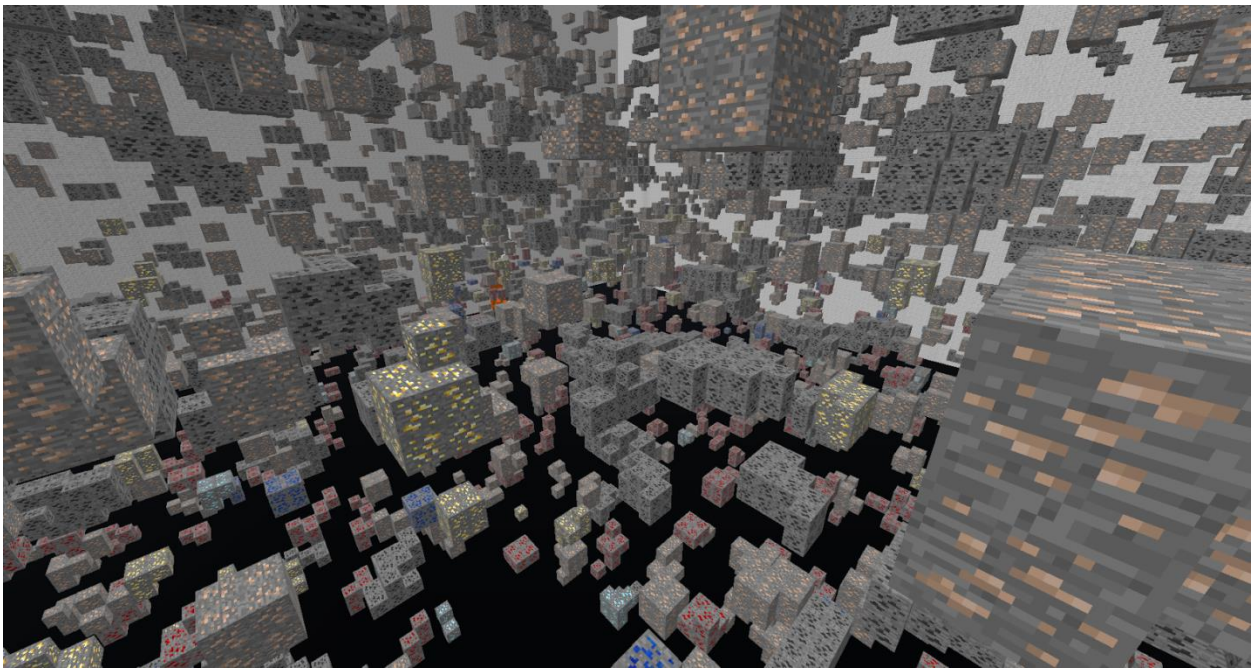
Ore Distribution In Minecraft



Alexander J. Schmidbauer

5/4/2017

I. Executive Summary

The popular voxel game Minecraft, which involves players gathering resources to build structures, includes a mining element. Players use pickaxes to mine underground and find ores. There are seven different ores; coal, iron, lapis lazuli, diamond, gold, redstone, and emeralds. Ores are distributed underground by an algorithm that takes into account depth and location. The map in Minecraft is divided into chunks, 16 x 16 block areas that extend from the bottom of the map to the sky. In every chunk, it is believed of the ores spawned, they will be 55% coal, 29.7% iron, 3.1% gold, 1.3% lapis lazuli, 9.7% redstone, and 1.2% diamond. Due to the rarity of emeralds, we exclude them from this study. This claim's validity was tested. Data was collected by observing an area of 8 x 8 chunks, or 64 chunks, and measuring the amount of the different ores found there. A one-way table was used in conjunction with a chi-square goodness of fit test to analyze the data found. The claim provided was proven to be untrue, as the data found can be made to provide enough evidence that the true ore distribution does not match the theoretical distribution given by the null hypothesis.

II. Introduction

The question provided in this study is: Are the ores underground in Minecraft distributed by chunk as 55% coal, 29.7% iron, 3.1% gold, 1.3% lapis lazuli, 9.7% redstone, and 1.2% diamonds per chunk? The variable we are interested in through this study is the type or ores, with coal, iron, lapis lazuli, redstone and diamond categories. The null hypothesis (H_0) states that the respective ore distribution is as follows: $p_1 = 0.55$, $p_2 = 0.297$, $p_3 = 0.031$, $p_4 = 0.013$, $p_5 = 0.097$, and $p_6 = 0.012$, of which $p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = 1$. The alternative hypothesis (H_1) states that the ores are not distributed this way, such that at least one of the ore distributions differs from the probabilities provided in the null hypothesis. This analysis will use a one-way table with a chi-square goodness of fit test at $\alpha = 0.5$.

III. Data Collection Techniques

For this study, an area of 8 x 8 chunks, or 64 chunks was chosen in Minecraft. The data was obtained by using external map editors to remove every block except those needed in the study from the region, and then counting the different ores found in the region. 64 chunks were chosen to study in order to reduce error in the analysis. The total amount of ores was also found to be n = 22373. The values and their associated percentages were recorded in a table.

IV. Summary Information

Table 1:

Ore One-Way Table							
Ore	Coal	Iron	Gold	Lapis Lazuli	Redstone	Diamond	Total
# Of Ore	11806	7283	770	291	1971	252	22373

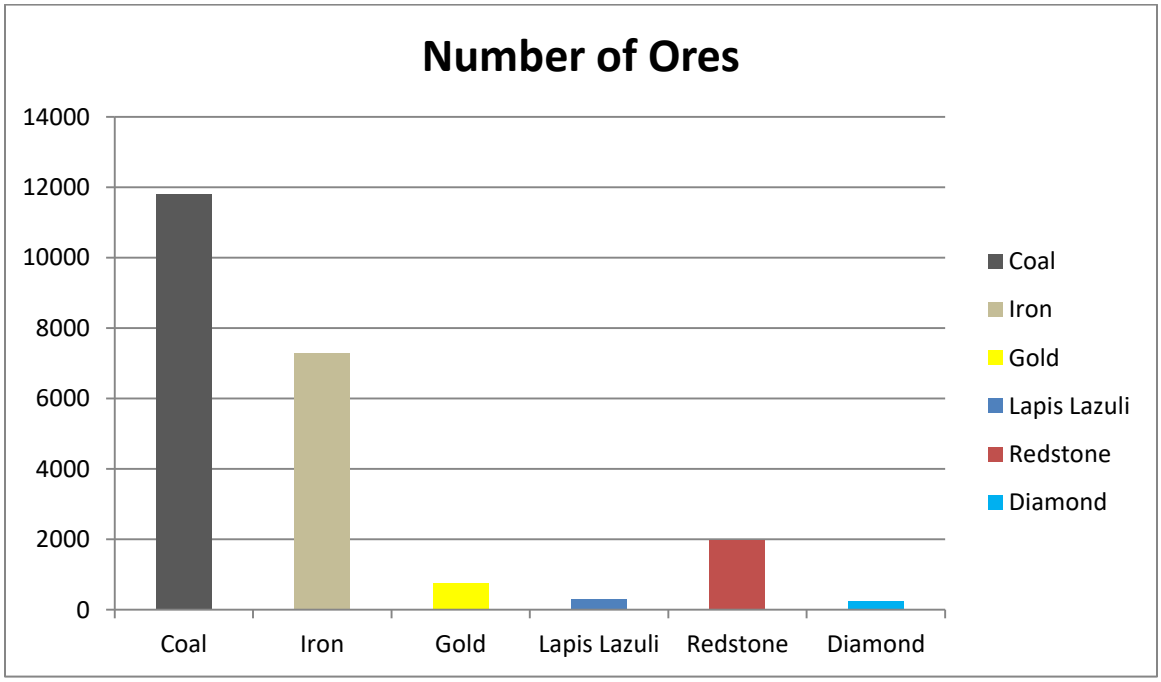
Table 1 shows ore distribution found in the 8 x 8 chunk area. The total found in the last row is considered the sample size.

Table 2:

Ore	Frequency	Percent	Cumulative Percent
Coal	11806	52.77	52.77
Iron	7283	32.55	85.32
Gold	770	3.44	88.76
Lapis Lazuli	291	1.31	90.07
Redstone	1971	8.81	98.88
Diamond	252	1.12	100.00
Total	22373	100.00	100.00

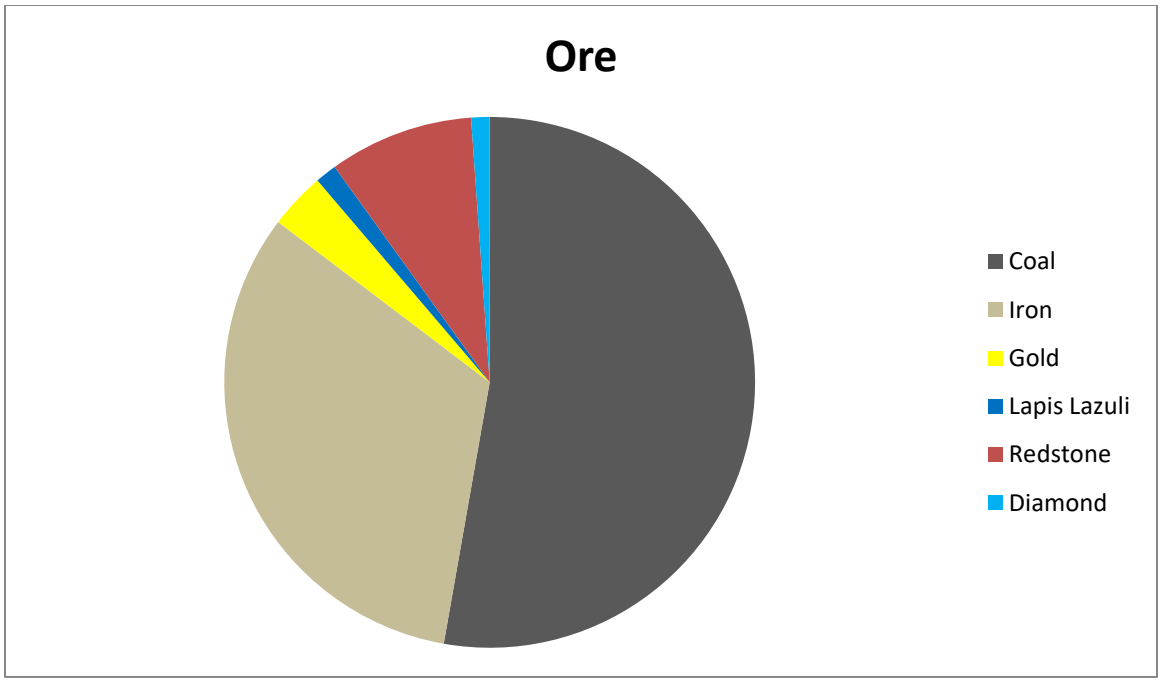
Table 2 shows frequency, percentage and cumulative percentage of each ore. These values were used to construct a bar graph and pie chart.

Graph 1: Bar Graph of Ore Distribution



Graph 1 shows the frequency of the ores found in the 8 x 8 chunk area.

Graph 2: Pie Chart of Ore Distribution



Graph 2 shows the percentage of each ore distributed observed.

V. Analysis

A chi-square goodness-of-fit test was used to analyze the data to see if the data matches the distribution provided. The hypothesis (H_0) tested states that the respective ore distribution is as follows: $p_1 = 0.55$, $p_2 = 0.297$, $p_3 = 0.031$, $p_4 = 0.013$, $p_5 = 0.097$, and $p_6 = 0.012$. The alternative hypothesis (H_1) states that the ores are not distributed this way, such that at least one of the ore distributions differs from the probabilities provided in the null hypothesis.

The test statistic used is $\chi^2 = \sum \frac{(n_i - E_i)^2}{E_i}$, where n_i is the observed frequency for the ore, and E_i is the expected frequency for each ore. The expected frequencies are listed in Table 3.

Table 3

Ore	Percent	Expected Frequency
Coal	55	12305
Iron	29.7	6644
Gold	3.1	694
Lapis Lazuli	1.3	291
Redstone	9.7	2170
Diamond	1.2	268

Table 4

Ore	Expected Frequency	Observer Frequency	$\frac{(n_i - E_i)^2}{E_i}$
Coal	12305	11806	20.23576
Iron	6644	7283	61.4571
Gold	694	770	8.322767
Lapis Lazuli	291	291	0
Redstone	2170	1971	18.24931
Diamond	268	252	0.955224
Total	22373	22373	109.2202

Table 4 shows the Expected Frequency in comparison to the Observed Frequency.

A rejection region of $\chi^2 > 11.0705$ is chosen based on a confidence interval of 95% at $\alpha = 0.05$ and 5 degrees of freedom. The test statistic is found to be 109.2202, well within the rejection region. The data provided proves the ore distribution is not what it was believed to be, with each ore type except lapis lazuli differing from the proportions given in the null hypothesis.

VI. Conclusion

The results obtained with the chi-square goodness-of-fit test provide evidence to reject the null hypothesis that the ores are distributed by chunk as 55% coal, 29.7% iron, 3.1% gold, 1.3% lapis lazuli, 9.7% redstone, and 1.2% diamonds per chunk The alternative hypothesis is strongly supported that the true proportion is different for each ore.

This study has a few limitations. Ore distribution is also based on the biome provided by the game, so it may be possible the averages are based on all biomes rather than the small sample used to collect data. Second, enough data may not have been collected due to processing limitations on the computer used. In addition, the generation of caves, mineshafts, dungeons and landforms on the map may also affect the amount of ores found, as they generate after the ores and may overwrite them on the map.

Additional research may be done by disabling certain features in the game and generating a larger sample size. Examination of the algorithm used for world generation may also reveal certain intricacies not known about.