Question 1.2.

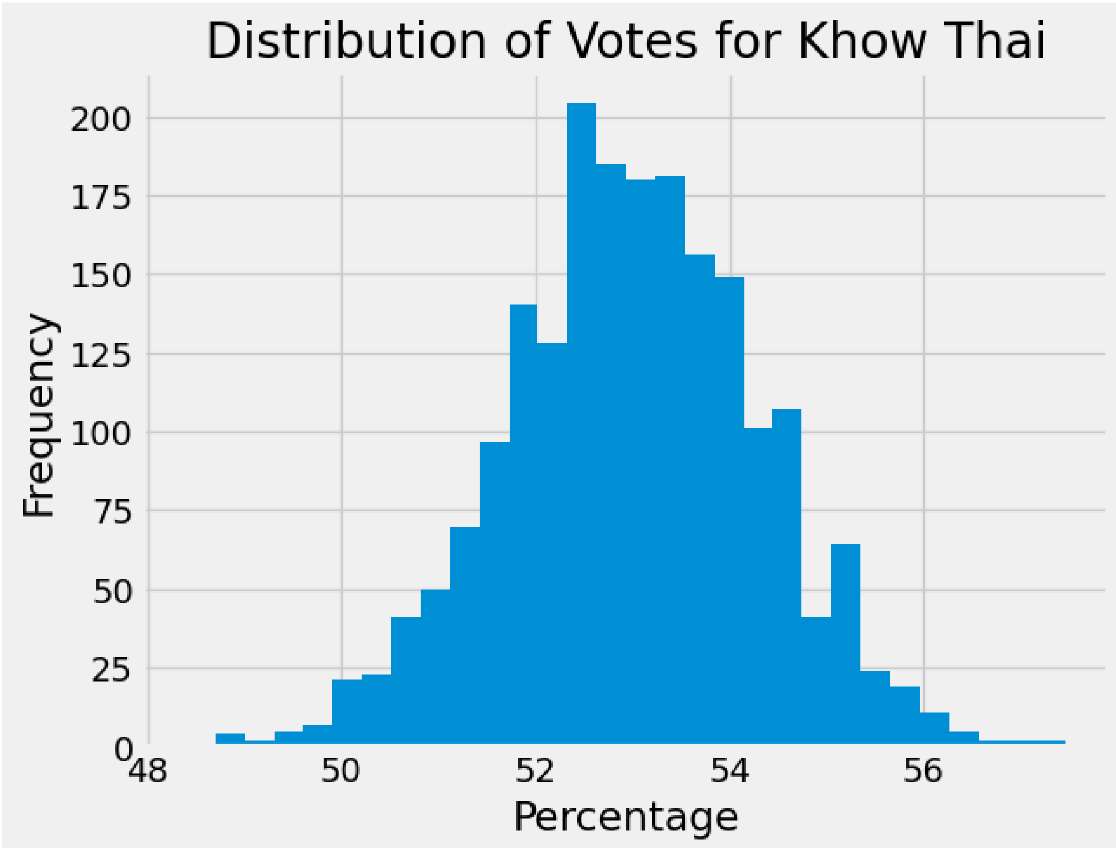
a). Complete the `percentages_in_resamples` function such that it simulates and returns a numpy array of 2022 elements, where each element represents a bootstrapped estimate of the percentage of voters who will vote for Khow Thai. You should use the `one_resampled_percentage` function you wrote above.

b). Then run your function `percentages_in_resamples` and store the results in a numpy array called `resamples_percentages`. Then create a density histogram of the entries in `resamples_percentages` array. Label your axes and include a title on your plot.

```
In [6]: def percentages_in_resamples():
            bootstraps = np.array([])
            for i in range(2022):
                bootstraps = np.append(bootstraps, one_resampled_percentage(votes))
            return bootstraps
```

```
In [7]: resampled_percentages = percentages_in_resamples()
        plt.hist(resampled_percentages, bins=30)
        # your code for the histogram above here.  The code below plots a red marker at the mean:
        plt.title("Distribution of Votes for Khow Thai")
        plt.ylabel("Frequency")
        plt.xlabel("Percentage")
        # your code for histogram above this line
```

```
Out[7]: Text(0.5, 0, 'Percentage')
```




```
In [8]: grader.check("q1_2")
```

```
Out[8]: q1_2 results: All test cases passed!
```

Question 2.5. The staff also created 70%, 90%, and 99% confidence intervals from the same sample, but we forgot to label which confidence interval represented which percentages! **First**, match each confidence level (70%, 90%, 99%) with its corresponding interval in the cell below

(e.g. ___ % CI: [52.1, 54] $\rightarrow$ replace the blank with one of the three confidence levels).

Then, explain your thought process and how you came up with your answers.

The intervals are below:

- [50.03, 55.94]
- [52.1, 54]
- [50.97, 54.99]

99% CI: [50.03, 55.94] 90% CI: [50.97, 54.99] 70% CI: [52.1, 54]

As the confidence interval decreases, the range will also decrease. As the confidence interval increases, the range will also increase.

Question 4.1. Michelle wants to use 10,000 bootstrap resamples to compute a confidence interval for the proportion of all Colorado voters who will vote Yes.

a). Use bootstrap resampling to simulate 10,000 election outcomes, and assign the np.array `resample_yes_proportions` to contain the Yes proportion of each bootstrap resample.

b). Calculate the 95% bootstrapped confidence interval for the Yes proportion.

c). Then, plot a density histogram of `resample_yes_proportions`. Include a title and label both axes. **You should see a bell shaped curve centered near the proportion of Yes in the original sample.** We have provided code that overlays your confidence interval at the bottom of your histogram.

In [42]:
```
def resampled():
    resample = np.random.binomial(400, .525)
    return resample/400

resample_yes_proportions = np.array([])
for i in range(10000):
    resample_yes_proportions = np.append(resample_yes_proportions, resampled())
# Your code to simulate empirical distribution above this line
resample_yes_proportions
```

Out[42]:
```
array([0.5475, 0.505 , 0.5375, …, 0.5975, 0.545 , 0.54  ])
```

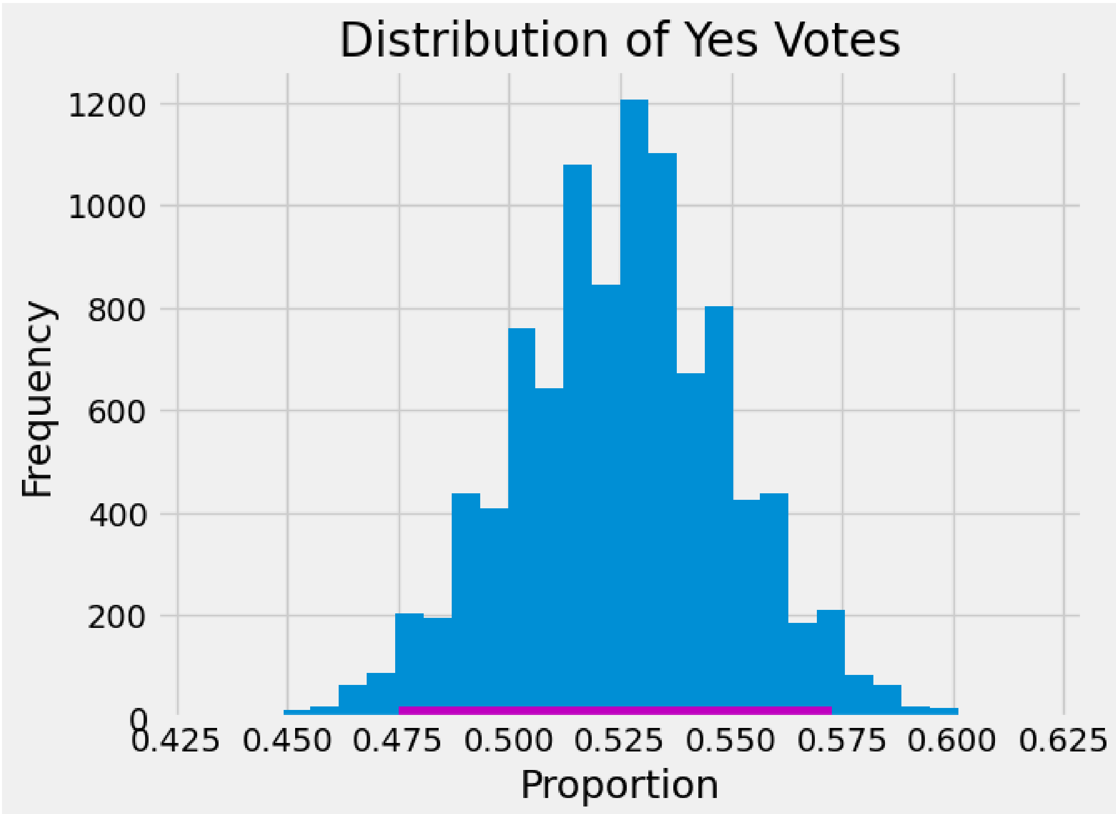
In [43]:
```
CI_lower = np.percentile(resample_yes_proportions,2.5)
CI_upper = np.percentile(resample_yes_proportions,97.5)
[CI_lower,CI_upper]
```

Out[43]:
```
[0.475, 0.5725]
```

In [45]:
```
plt.hist(resample_yes_proportions, bins=30)
plt.title("Distribution of Yes Votes")
plt.ylabel("Frequency")
plt.xlabel("Proportion")
# Your code to plot histogram above this line

plt.plot(np.array([CI_lower, CI_upper]), np.array([0, 0]), c='m', lw=10)
```

Out[45]:
```
[<matplotlib.lines.Line2D at 0x7fb34042a7a0>]
```




```
In [68]: grader.check("q4_1")
```

```
Out[68]: q4_1 results: All test cases passed!
```

Question 4.2.

a). Why does the Central Limit Theorem (CLT) apply in this situation, and how does it explain the distribution we see above?

b). Prove the following: In a population whose members are 0 or 1, the **standard deviation** of that population is:

$$\text{standard deviation of population} = \sqrt{(\text{proportion of 0s}) \times (\text{proportion of 1s})}$$

Write up your answers to both parts in the same Markdown cell below:

a). The Central Limit Theorem applies to this situation because the proportions of yeses involves summing our large random sample drawn with replacement. Furthermore, the CLT explains that the distribution of our random variable will be normal. Thus, the distribution we see above is normal.

b).

$SD[X] = \sqrt{Var[X]}$

$Var[X] = E[X^2] + (E[X])^2$

$E[X^2] = 1^2 * p + 0 * (1 - p) = p$

$E[X] = 1 * p + 0 * (1 - p) = p$

$(E[X])^2 = p^2$

$Var[X] = p - p^2 = p(1 - p)$

$SD[X] = \sqrt{p(1 - p)}$