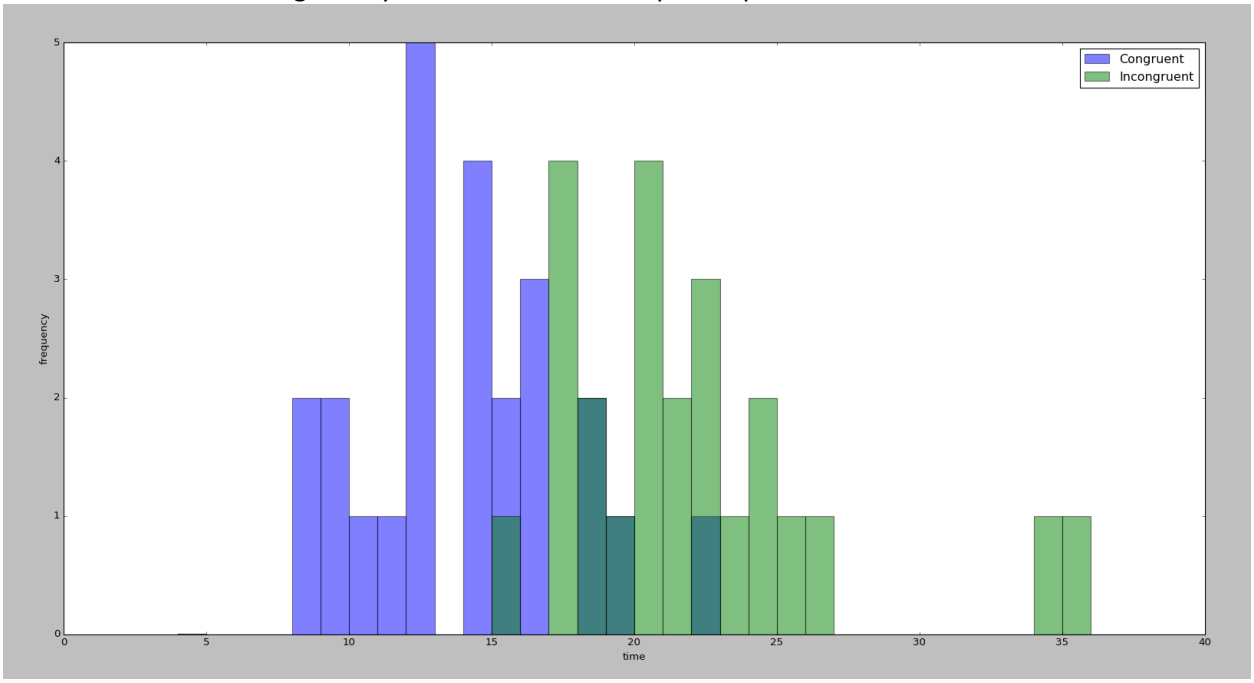
1. What is our independent variable? What is our dependent variable?
    a. Independent variable: The word condition of the list i.e. congruent or incongruent
    b. Dependent variable: The time it takes to name the ink colors in equally-sized lists
2. What is an appropriate set of hypotheses for this task? What kind of statistical test do you expect to perform? Justify your choices.
    a. Null hypothesis: $\mu_c \geq \mu_i$, where μ_c is the average time for a person to go through a list of congruent words and μ_i is the average time for a person to go through a list of incongruent words. The null says being congruent has no effect, or even will cause people to take longer time on average to read than being incongruent. I chose this so we will be doing a one tailed test since I suspect incongruent words only take a longer time to read, but not shorter.
    Alternative hypothesis: $\mu_c < \mu_i$. Congruent word lists takes shorter time on average to read than incongruent word lists.
    b. Since the population parameters (mean and standard deviation) for reading a list won't be available to us, we will be performing a t-test. The hypothesis we have indicates a one-tailed test is more appropriate. Since we have the same group of people reading over the two lists, we have dependent samples. A 5% significance level is chosen for my experiment, which is a common choice.
3. Report some descriptive statistics regarding this dataset. Include at least one measure of central tendency and at least one measure of variability.

Statistics	Congruent	Incongruent
Count	24	24
Mean	14.05	22.02
Median	14.36	21.02
Q1	11.53	18.67
Q3	16.59	24.37
Sample Variance	12.67	23.01
Sample Standard Deviation	3.56	4.80

4. Provide one or two visualizations that show the distribution of the sample data. Write one or two sentences noting what you observe about the plot or plots.



- I have done a histogram with both congruent and incongruent columns from the data plotted on one graph. From the histogram, we can see that the distribution for the time to read off a congruent and an incongruent list are both positively skewed. The variance in time to read off a list is smaller for congruent lists. The same can be said for the average time.

5. Now, perform the statistical test and report your results. What is your confidence level and your critical statistic value? Do you reject the null hypothesis or fail to reject it? Come to a conclusion in terms of the experiment task. Did the results match up with your expectations?
Here, we have a dependent sample so we can use the formula $\mathrm{SE} = \mathrm{s_{diff}}/\sqrt{n}$ to find our sample deviation. The null can be rewritten as $\mu_i - \mu_c \geq 0$. Hence we define the difference as $\mu_i - \mu_c$.

Sample Size	24
Standard Error	0.99
Average Difference	7.96
t-statistic	8.02
Confidence level	0.95
Significance level	0.05
Critical value	1.714
p-value	0.00000002

We perform a one tailed t-test here at confidence level of 95%. At this confidence level, as p<0.05 and t-statistic is greater than the critical value, the statistical test shows there is sufficient evidence to reject the null hypothesis. This means that in general, it takes people longer on average to read off a list of incongruent words.
This result match up to my expectations.

6. Optional: What do you think is responsible for the effects observed? Can you think of an alternative or similar task that would result in a similar effect? Some research about the problem will be helpful for thinking about these two questions!
    a. It takes the brain longer to process color than words.
    b. The automaticity theory suggests that recognizing colors is not an "automatic process" there is hesitancy to respond; whereas, the brain automatically understands the meaning of words as a result of habitual reading.
    c. Similar tasks: print words “small medium large” in small medium and large fonts and ask people to read out the font size from a congruent list and an incongruent list.

Code used to produce plot in question 4

```
plt.hist(data['Congruent'],bins,alpha=0.5,label='Congruent')
plt.hist(data['Incongruent'],bins,alpha=0.5,label='Incongruent')
plt.legend(loc='upper right')
plt.xlabel('time')
plt.ylabel('frequency')
plt.show()
```