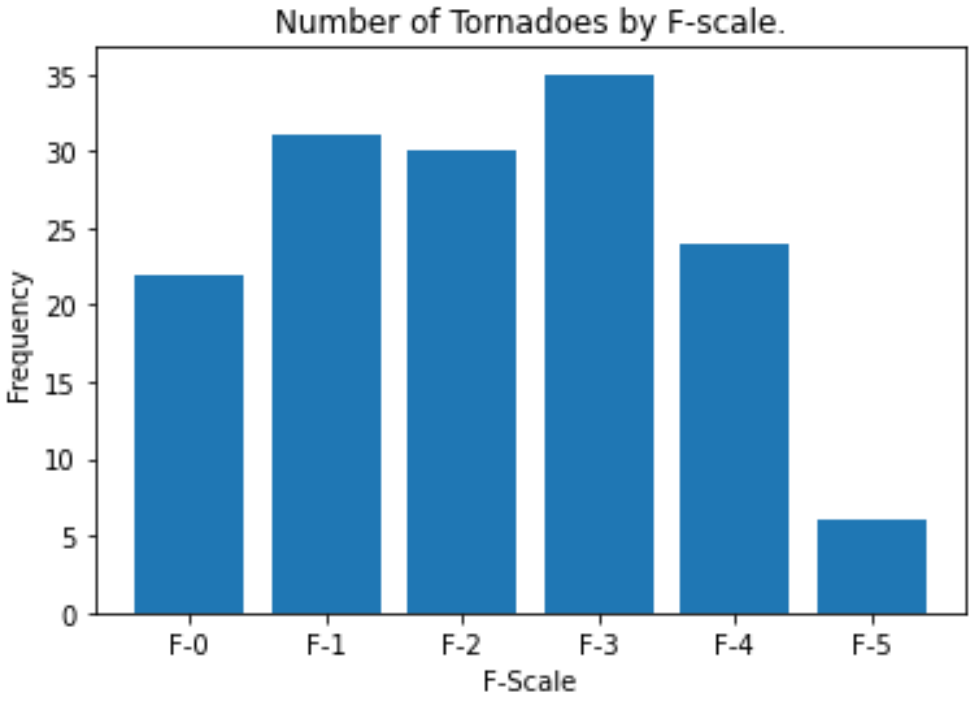
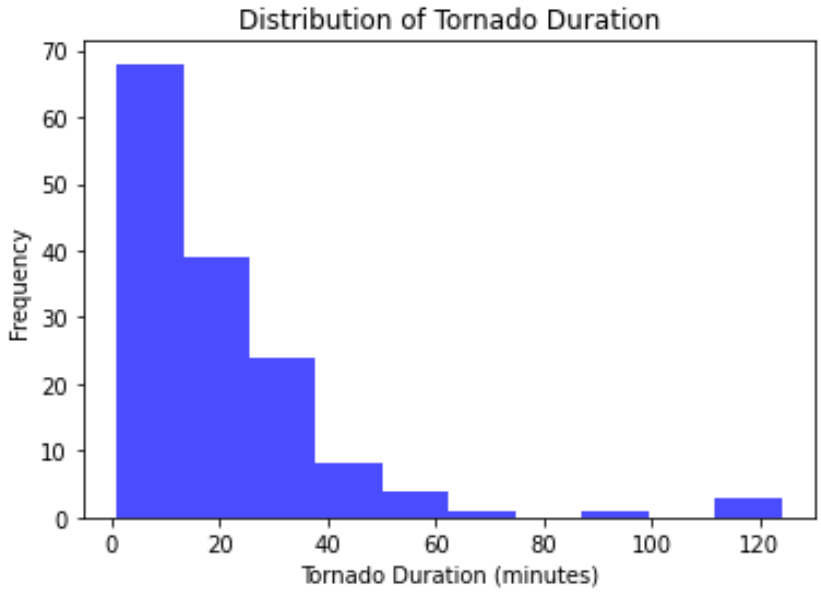
Assignment 1: Graphical representation of Data
Lincoln Nordquist

1. The bar graph indicates that roughly 20% of the tornadoes exceeded an f-scale of 3. This is very much inconsistent with typical reports since generally only 1% of tornadoes exceed an F-scale of 3. This means that the tornados in this particular case study were much deadlier than normal.



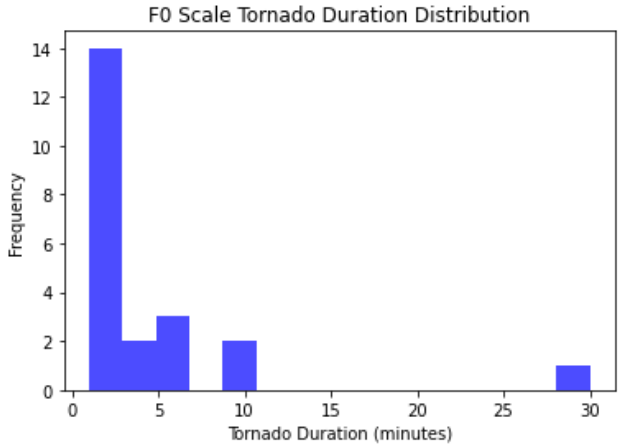

```python
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

data = pd.read_csv("assignment_1_data.csv")
d = pd.DataFrame(data)


# DURATION CONSTANTS
F0_DURATION = [1, 1, 5, 1, 1, 6, 4, 10, 5, 4, 1, 1, 1, 1, 1, 1, 1, 1, 30, 1, 9]
F1_DURATION = [16, 13, 9, 8, 13, 10, 15, 1, 17, 23, 10, 8, 12, 5, 20, 31, 12, 5, 30, 13, 7, 1, 5, 13, 1, 2, 5, 10, 1, 20, 5]
F2_DURATION = [7, 15, 2, 10, 23, 10, 7, 12, 8, 1, 8, 19, 5, 10, 15, 20, 10, 13, 20, 15, 13, 14, 1, 4, 2, 15, 30, 91, 11, 5]
F3_DURATION = [9, 20, 8, 16, 26, 36, 10, 20, 50, 17, 26, 31, 21, 30, 23, 28, 23, 18, 35, 35, 15, 25, 30, 15, 22, 18, 58, 19, 23, 31, 13, 26, 40, 14, 11]
F4_DURATION = [120, 23, 23, 42, 47, 25, 22, 22, 34, 50, 38, 28, 39, 29, 28, 25, 34, 16, 40, 55, 124, 30, 30, 31]
F5_DURATION = [37, 69, 23, 52, 61, 122]

# CODE FOR QUESTION 1
frequency = np.array([len(F0_DURATION), len(F1_DURATION), len(F2_DURATION), len(F3_DURATION), len(F4_DURATION), len(F5_DURATION)])
my = ["F-0", "F-1", "F-2", "F-3", "F-4", "F-5"]
plt.pie(frequency, labels = my)
plt.show()
plt.xlabel('F-Scale')
plt.ylabel('Frequency')
plt.title('Number of Tornadoes by F-scale.')

print(plt.bar(my,frequency))
plt.show()
```

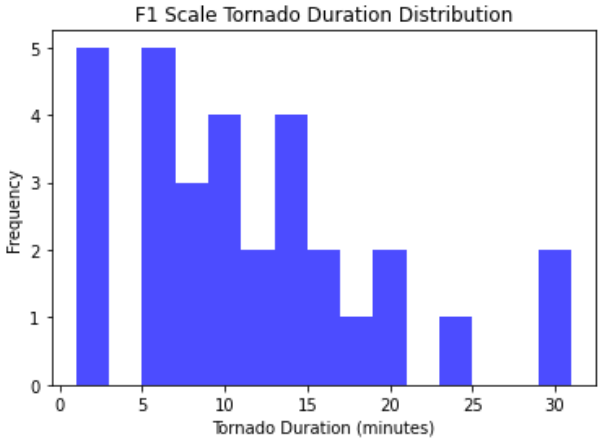
2. This histogram represents the frequency of tornadoes based on duration. The histogram shows that the shorter tornados happened much more frequently than the longer ones. As you can see, the tornadoes that lasted between 0 and 20 seconds happened the most by far, and as the tornados decrease in length, the frequency decreases as well. The only surprising detail is that the frequency of tornadoes lasting 110-130 minutes is higher than the frequency of tornadoes lasting 90-100 minutes.



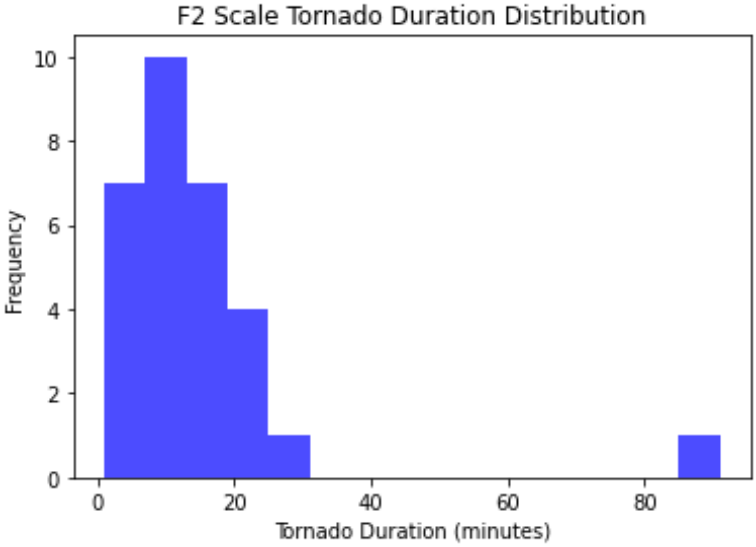

```
# CODE FOR QUESTION 2
histogram_data = F0_DURATION + F1_DURATION + F2_DURATION + F3_DURATION + F4_DURATION + F5_DURATION

plt.hist(histogram_data, bins=10, color='blue', alpha=0.7)
plt.xlabel('Tornado Duration (minutes)')
plt.ylabel('Frequency')
plt.title('Distribution of Tornado Duration')

plt.show()
```

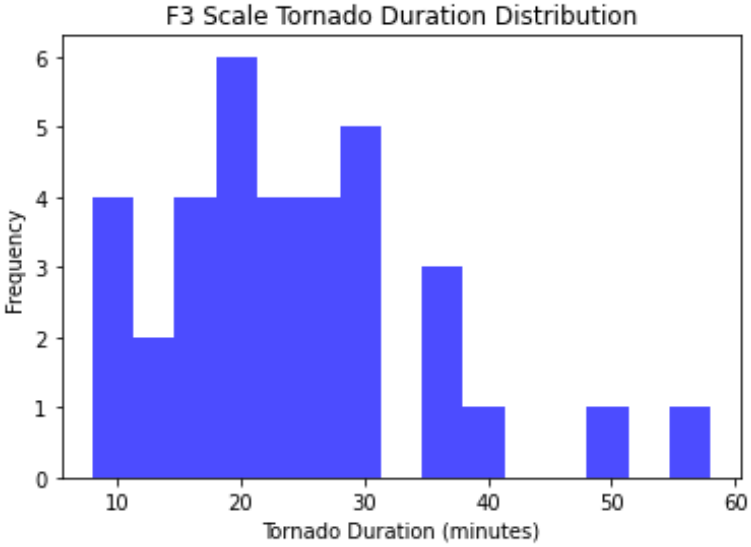
3. As we can see in the histograms of F0 and F1, there definitely seems to be a higher frequency of tornadoes with a shorter duration. This is very apparent in the F0 histogram since there are over 15 tornadoes that last less than 5 minutes. The F1 histogram definitely follows more of a trend though, and with a much larger sample size, we can clearly see that as the tornadoes increased in length, the frequency of them decreased as well.



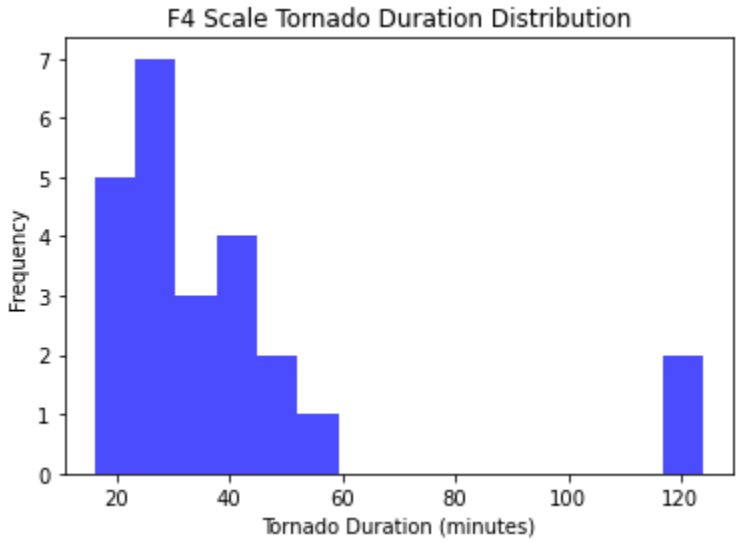


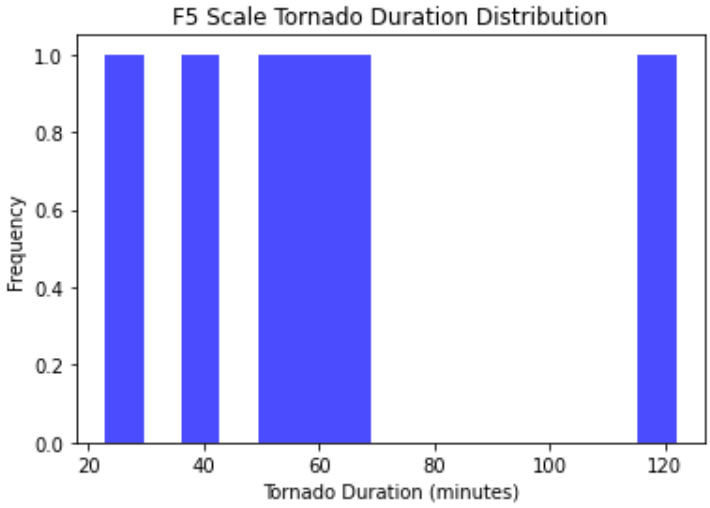
In histograms for F2 and F3, we once again see a higher frequency of tornadoes taking place in shorter duration, but there appears to be a little bit more variation as well. In the F2 histogram we see the tornadoes spiking at approximately 10 minutes, which is inconsistent with the previous data. My hypothesis is that this is likely due to too small of a sample size, which can skew the data.





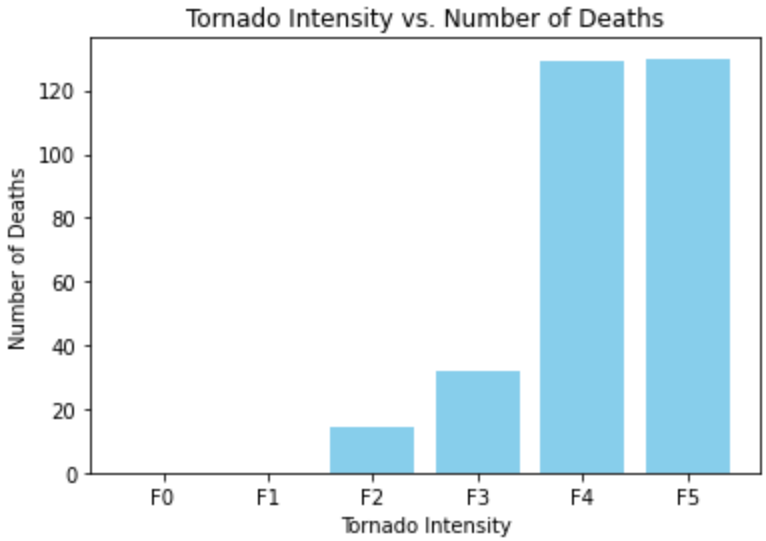
Finally, in the F4 and F5 histograms, we unfortunately have a smaller sample size to work with, but we are still able to interpret what we have. The F4 histogram is very similar to the F2 one, because there is a similar spike in frequency, which is inconsistent with the correlation. In the F5 histogram, we can clearly see that there is a much higher frequency of 20-70 minute tornadoes than there are 120 minute tornadoes, which is once again, consistent with our previous histograms.





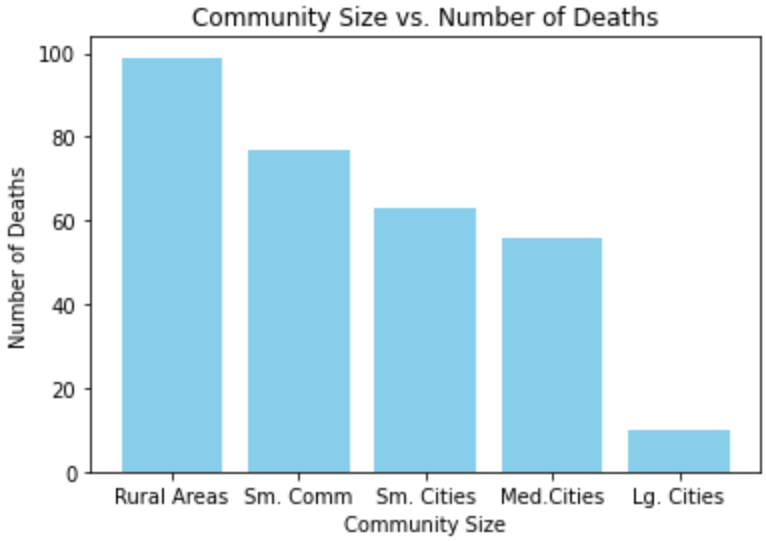
```
# CODE FOR QUESTION 3
# F0 histogram
f0_data = F0_DURATION

plt.hist(f0_data, bins=15, color='blue', alpha=0.7)
plt.xlabel('Tornado Duration (minutes)')
plt.ylabel('Frequency')
plt.title('F0 Scale Tornado Duration Distribution')

plt.show()

# F1 histogram
f1_data = F1_DURATION

plt.hist(f1_data, bins=15, color='blue', alpha=0.7)
plt.xlabel('Tornado Duration (minutes)')
plt.ylabel('Frequency')
plt.title('F1 Scale Tornado Duration Distribution')

plt.show()

# F2 histogram
f2_data = F2_DURATION

plt.hist(f2_data, bins=15, color='blue', alpha=0.7)
plt.xlabel('Tornado Duration (minutes)')
plt.ylabel('Frequency')
plt.title('F2 Scale Tornado Duration Distribution')

plt.show()

# F3 histogram
f3_data = F3_DURATION

plt.hist(f3_data, bins=15, color='blue', alpha=0.7)
plt.xlabel('Tornado Duration (minutes)')
plt.ylabel('Frequency')
plt.title('F3 Scale Tornado Duration Distribution')

plt.show()

# F4 histogram
f4_data = F4_DURATION

plt.hist(f4_data, bins=15, color='blue', alpha=0.7)
plt.xlabel('Tornado Duration (minutes)')
plt.ylabel('Frequency')
plt.title('F4 Scale Tornado Duration Distribution')

plt.show()

# F5 histogram
f5_data = F5_DURATION

plt.hist(f5_data, bins=15, color='blue', alpha=0.7)
plt.xlabel('Tornado Duration (minutes)')
plt.ylabel('Frequency')
plt.title('F5 Scale Tornado Duration Distribution')

plt.show()
```

4. This next bar chart represents the relationship between tornado intensity and deaths. The case study states that the most severe tornadoes (F4 and F5) account for over 70% of deaths. This bar chart clearly aligns with those numbers as we can see here. The F4 and F5 tornadoes each account for roughly 120 deaths each, while the F2 and F3 tornadoes only account for roughly 15 and 30 deaths. Meanwhile, the F0 and F1 tornadoes weren't responsible for any deaths at all. This bar chart shows a clear positive correlation between the intensity of a tornado and the number of deaths it causes.




```
# CODE FOR QUESTION 4
F0_DEATHS = 0
F1_DEATHS = 0
F2_DEATHS = 14
F3_DEATHS = 32
F4_DEATHS = 129
F5_DEATHS = 130

tornado_intensity = ["F0", "F1", "F2", "F3", "F4", "F5"]
number_of_deaths = [F0_DEATHS, F1_DEATHS, F2_DEATHS, F3_DEATHS, F4_DEATHS, F5_DEATHS]

plt.bar(tornado_intensity, number_of_deaths, color='skyblue')
plt.xlabel('Tornado Intensity')
plt.ylabel('Number of Deaths')
plt.title('Tornado Intensity vs. Number of Deaths')

plt.show()
```

5. The last histogram shows the relationship between community size and number of deaths. This bar chart shows a clear pattern, which is that the larger the community, the less deaths that are caused by tornadoes. One inference that could be made is that less tornadoes happen in larger communities. Another inference could be that perhaps larger communities have safer and more sturdy infrastructure than the smaller communities.




```
117
118  # CODE FOR QUESTION 5
119
120  RURAL_AREAS = 99
121  SMALL_COMMUNITIES = 77
122  SMALL_CITIES = 63
123  MEDIUM_CITIES = 56
124  LARGE_CITIES = 10
125
126  community_size = ["Rural Areas", "Sm. Comm", "Sm. Cities", "Med.Cities", "Lg. Cities"]
127  number_of_deaths = [RURAL_AREAS, SMALL_COMMUNITIES, SMALL_CITIES, MEDIUM_CITIES, LARGE_CITIES]
128
129  plt.bar(community_size, number_of_deaths, color='skyblue')
130  plt.xlabel('Community Size')
131  plt.ylabel('Number of Deaths')
132  plt.title('Community Size vs. Number of Deaths')
133
134  plt.show()
135
```

There were several interesting things I found while analyzing the data from the case study. One is that this particular case study was very much unlike most recorded tornadoes. This case study consisted of a much greater frequency of F4 and F5 scale tornadoes. Despite this, there seemed to be a very consistent correlation between the frequency of a tornado, and its duration. Across all six tornado intensity, the ones that had the shortest duration, had the highest frequency. I found it very interesting that despite the intensities of the tornadoes differing by a lot, they still maintained the same correlation.

Another thing I found, which didn't surprise me, was that there was a significantly greater number of deaths to F4 and F5 tornadoes than F1-F3. This was unsurprising since it makes sense that more violent tornadoes would cause more deaths than less violent ones. This made me come to the conclusion that the more violent a tornado is, the more deaths it will cause, and the less frequent it will occur.

The final correlation I was able to determine was that the smaller a community is, the more deaths will be caused by tornadoes. I was able to come to the conclusion that deaths happen as a function of community size. This means that as the community size shrinks, the deaths will proportionally increase along with it.