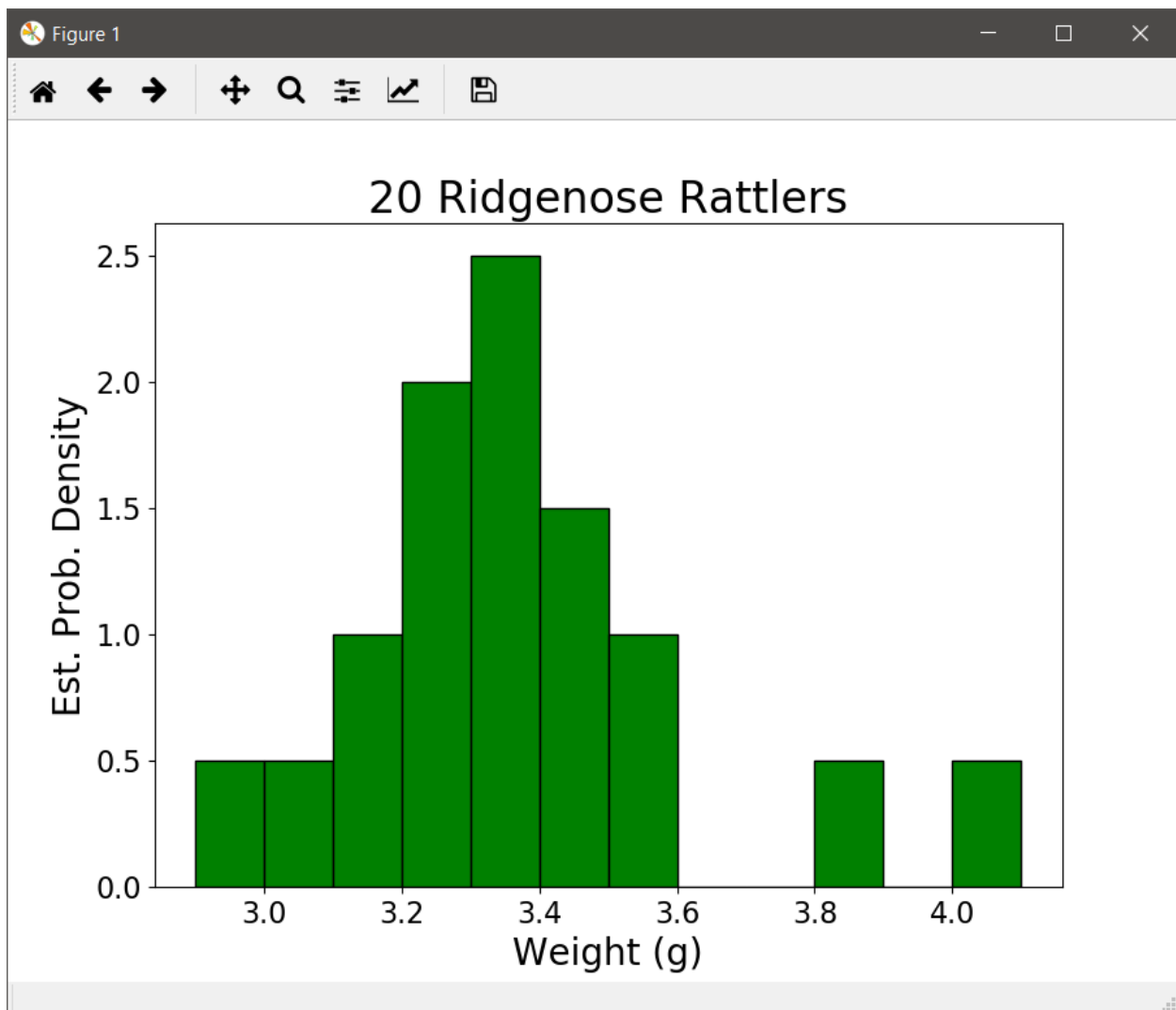


ISTA 350 Lab 8 Worksheet, 3/30/2020**Name:**

Draw a histogram representing the estimated probability density function for the weights of adult Ridgenose Rattlesnakes in grams using the following sample: 3.24, 4.01, 3.08, 3.56, 3.34, 3.43, 2.98, 3.81, 3.29, 3.52, 3.47, 3.38, 3.31, 3.19, 3.21, 3.35, 3.17, 3.22, 3.31, 3.41. Make your bins to be width 0.1 with bin edges lying on 2.9, 3.0, 3.1... What is the estimated probability that a Ridgenose Rattlesnake will weigh between 3.3 and 3.4 grams? Between 3.3 and 3.5 grams? Greater than 3.3 grams? Less than 3.3 grams?



$$P(3.3 \leq wt \leq 3.4) = 2.5 * 0.1 = 0.25$$

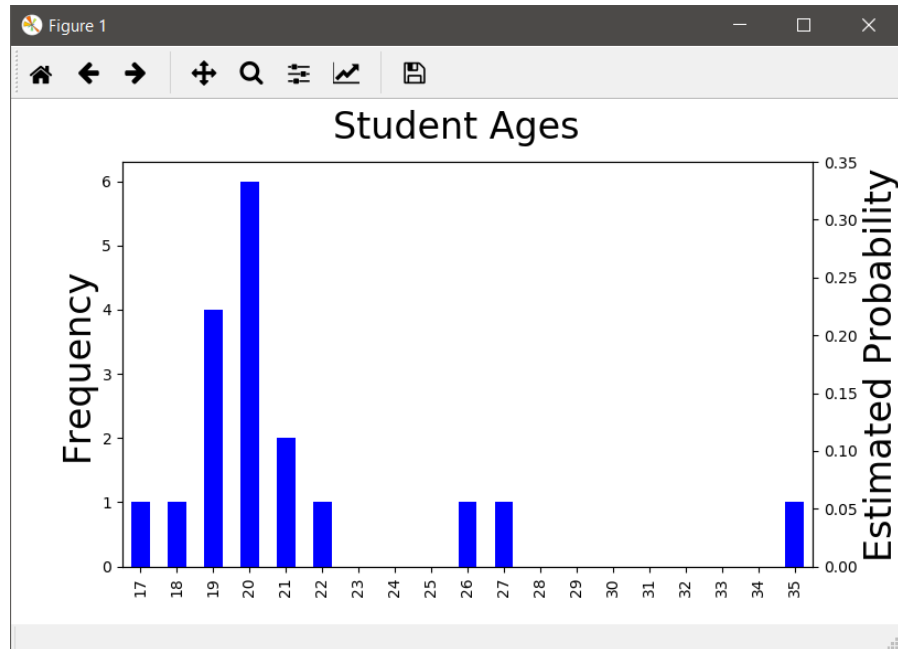
$$P(3.3 \leq wt \leq 3.5) = 2.5 * 0.1 + 1.5 * 0.1 = 0.40$$

Do less then first, it's easier:

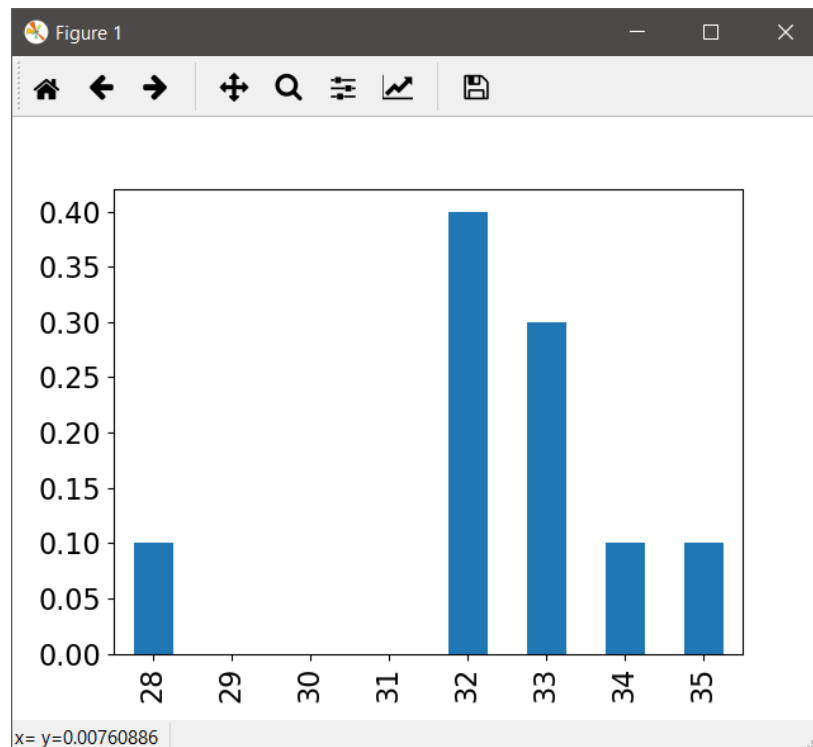
$$P(wt < 3.3) = (0.5 + 0.5 + 1.0 + 2.0) * 0.1 = 0.40$$

$$P(wt > 3.3) = 1 - P(wt < 3.3) = 0.60$$

Draw a bar chart representing the frequency distribution of the following values representing the ages of the students in a hypothetical UA class: 19, 20, 21, 17, 20, 19, 20, 22, 19, 18, 21, 20, 26, 35, 27, 19, 20, 20. On the right-hand side of your chart, add an axis and label it so that the bar chart represents a probability mass function (probability distribution).



Create a bar plot of the probability mass function for the following sequence of values: [33, 35, 28, 32, 33, 32, 34, 33, 32].



Write a few lines of code that create a `Series` which has the integers between 28 and 35 inclusive as its index and their associated probabilities as determined above as its data.

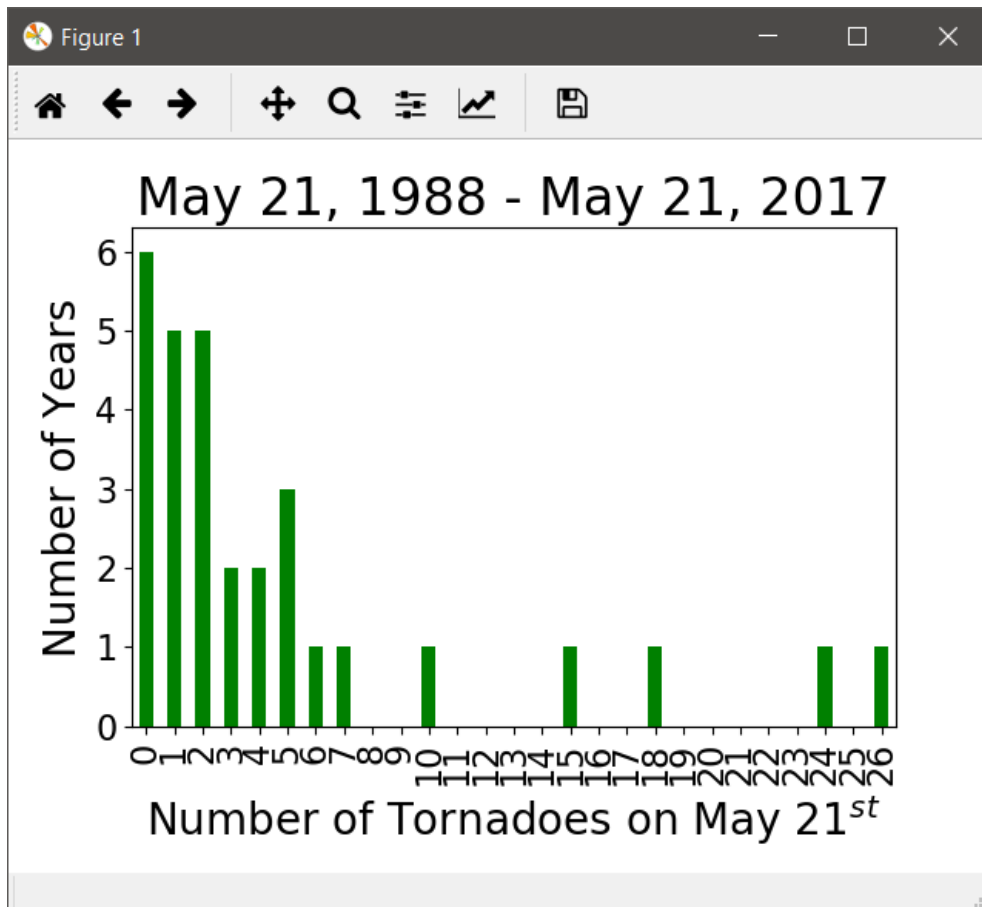
```
data = [33, 35, 28, 32, 33, 32, 32, 34, 33, 32]
s = pd.Series(data).value_counts().sort_index()
s = s.reindex(range(s.index[0], s.index[-1] + 1)).fillna(0)
s = s / s.sum()
```

Write a function called `frequency_series` that takes a list of values as its sole argument. It returns a `Series` whose `index` is the values in the list without duplicates and whose values are the corresponding frequencies.

```
def frequency_series(data):
    return pd.Series(data).value_counts()
```

Write a function called `frequencies` that takes a list of values as its sole argument. It returns two lists: the first list has the same values as the argument in the same order, but with no duplicates. The second list is the same length and each element is the number of times the corresponding element in the first list appears in the argument. For example, if the argument is `[99, 77, 99, 77, 77]`, return `[99, 77], [2, 3]`. There is a list method called `count` that does what it sounds like.

```
def frequencies(data):
    values = []
    counts = []
    for i in range(len(data)):
        if data[i] not in values:
            values.append(data[i])
            counts.append(data.count(data[i]))
    return values, counts
```



What is the estimated probability that there will be a tornado on May 21st of 2019? (Hints: What is the probability that there will NOT be a tornado? You will need to turn these frequencies into probabilities. What is the number of observations, N ?) What is our estimated probability that there will be a double-digit number of tornadoes? What is our estimated probability that there will be one or two tornadoes?

To turn our frequencies into estimated probabilities, we need to divide by the number of observations, N . In this case $N = 30$ (we have 30 years of tornado counts for May 21st). The probability that there will be a tornado is $1 -$ the probability that there will not be a tornado. We had 6 years where there was no tornado on 5/21, so our estimated probability of no tornadoes is $6 / 30 = 0.2$, so our estimated probability that there will be a tornado is 0.8.

There were 5 years with double-digit tornadoes on 5/21, so our estimated probability of double-digit tornadoes is $5 / 30 = 0.167$.

There were 5 years with one tornado and 5 with two, so our probability of one or two tornadoes is $(5 + 5) / 30 = 0.333$. Notice that we calculated our probabilities by adding our frequencies and then dividing by 30. We could have calculated our probabilities first and then added them, i.e. $0.167 + 0.167 = 0.333$.