cheg304 homework 1 question 1

AUTHOR
anon for grading??

PUBLISHED
February 11, 2025

First, read in the data downloaded from canvas. then save the yield information for both catalyst A and catalyst B to np.arrays for easy use and also efficient storage (even though this data is tiny)

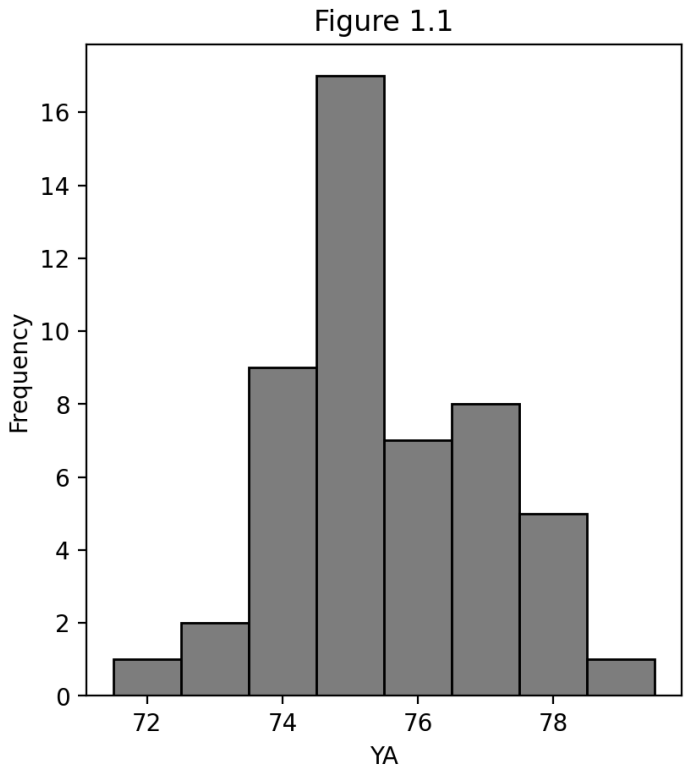
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_excel('Chapter1.xlsx', sheet_name='Table1.1')
YA = data['YA'].values
YB = data['YB'].values
```

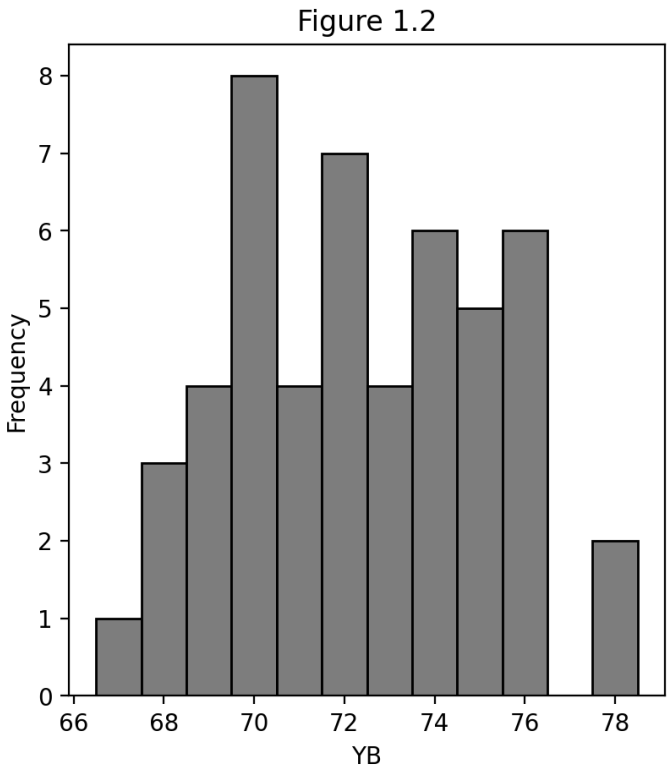
Now recreating 1.1 and 1.2

- the bins are centered above integers and width of 1%
- 1.1 has 8 bins
- 1.2 has 11 bins
- used same start and stop as book's figures

it said to carefully describe how it was plotted so... i started by making a figure and two axes(s?) with the `plt.subplots` function with 1 row and 2 columns. i then plotted histograms in each using `np.arange` to make the bins start and stop where i'd like them to and have width 1. everything else is aesthetics

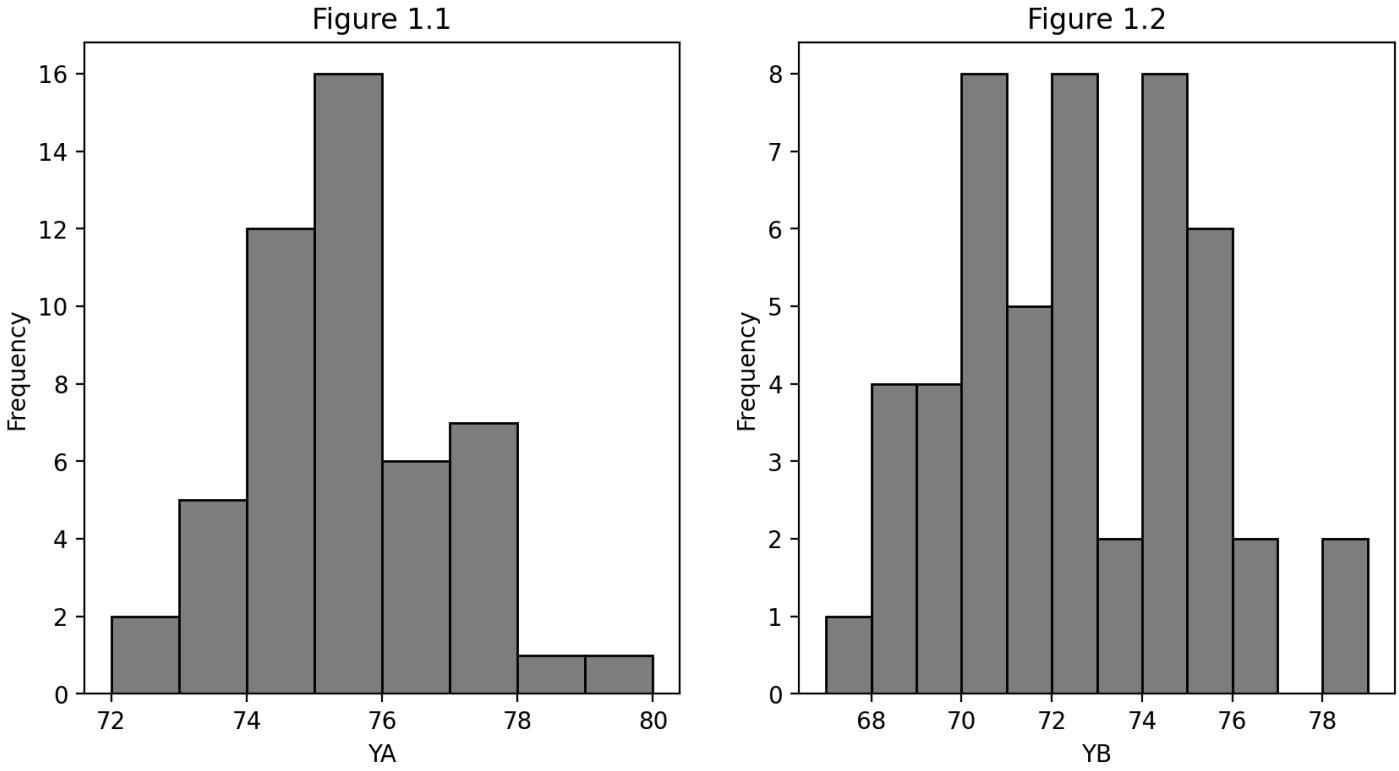
```
fig, (ax11, ax12) = plt.subplots(1,2, figsize=(10,5), dpi=200)

ax11.hist(YA, bins=np.arange(71.5,80.5,1), facecolor='grey', edgecolor='black')
ax12.hist(YB, bins=np.arange(66.5,79.5,1), facecolor='grey', edgecolor='black')

ax11.set(xlabel='YA', ylabel='Frequency', title='Figure 1.1')
ax12.set(xlabel='YB', ylabel='Frequency', title='Figure 1.2');
```





now moving the bins to have different midpoints. for plotting I followed the same as before but made the bins start on the integers (instead of $n + 0.5$) so that the bins would be centered between integers !!

```
fig, (ax11, ax12) = plt.subplots(1,2, figsize=(10,5), dpi=200)

ax11.hist(YA, bins=np.arange(72,81,1), facecolor='grey', edgecolor='black')
ax12.hist(YB, bins=np.arange(67,80,1), facecolor='grey', edgecolor='black')

ax11.set(xlabel='YA', ylabel='Frequency', title='Figure 1.1')
ax12.set(xlabel='YB', ylabel='Frequency', title='Figure 1.2');
```



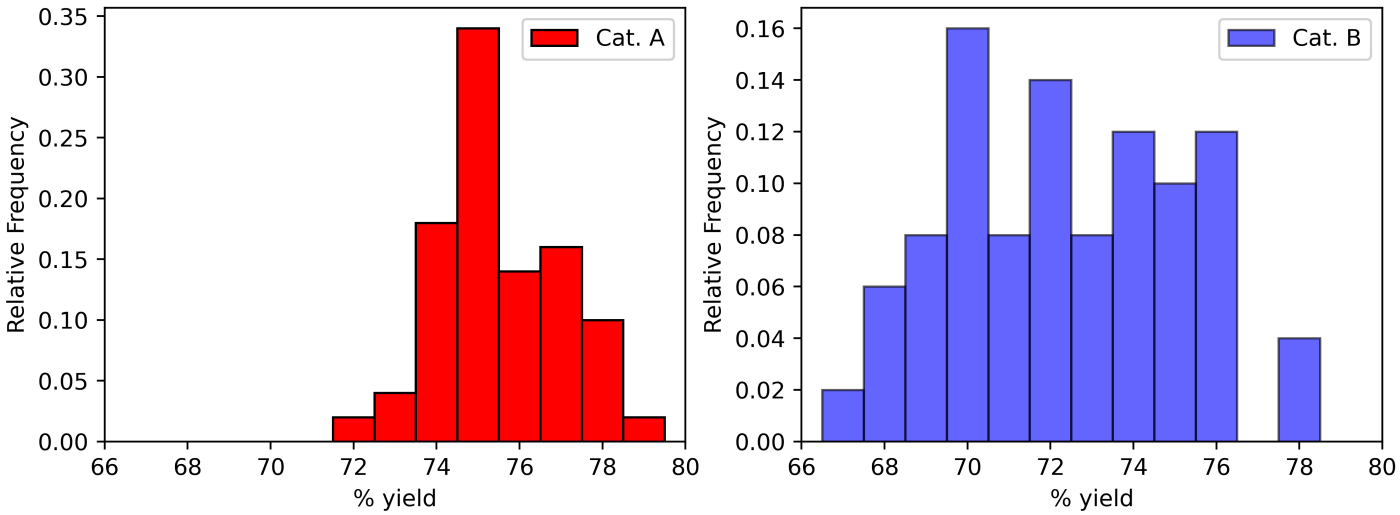
i feel that the representation of the data has not changed significantly–they look effectively the same.

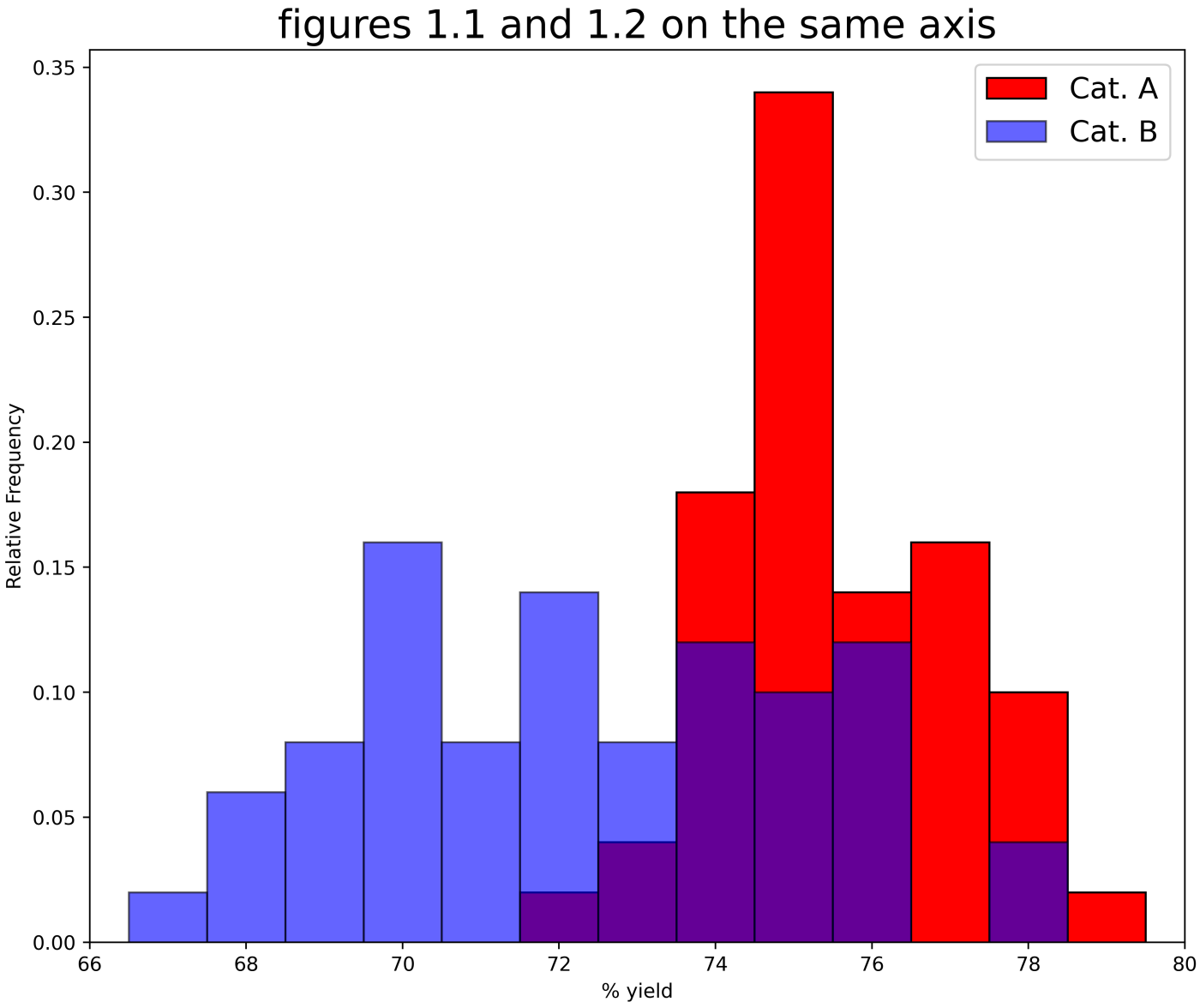
c.

to plot them together we need some coloring (see legend) and we need a way to see through the overlap. this could be done with 3d plots (fancy) or with side by side bars (ugly, disgusting, horrendous, torture) but i chose to just make the one rendered last be opaque.

```
fig, ax = plt.subplot_mosaic(
    [['1.1', '1.2'],
     ['both', 'both'],
     ['both', 'both']],
    figsize=(10,13),
    dpi=500,
    subplot_kw={'xlabel':'% yield', 'ylabel':'Relative Frequency', 'xlim':(66,80)}
)

fig.subplots_adjust(hspace=0.4)

ax['1.1'].hist(YA, bins=np.arange(71.5,80.5,1), facecolor='red', edgecolor='black', label='Cat.
ax['1.1'].legend()
ax['1.2'].hist(YB, bins=np.arange(66.5,79.5,1), facecolor='blue', edgecolor='black', alpha=0.6,
ax['1.2'].legend()


ax['both'].hist(YA, bins=np.arange(71.5,80.5,1), facecolor='red', edgecolor='black', label='Cat.
```

```
ax['both'].hist(YB, bins=np.arange(66.5,79.5,1), facecolor='blue', edgecolor='black', alpha=0.6,

ax['both'].set_title('figures 1.1 and 1.2 on the same axis', fontsize=20)
ax['both'].legend(fontsize=15);
```





conclusions made by the book:

1. A's 'center of action' is higher than B's
2. the values of B are more spread out than those of A
3. better chance of getting yield > 74.5 with catalyst a
4. it is unlikely that process a will ever return yields less than 71.50

5. the chance that process B will have yield < 71.50 is non-insignificant

(page 20, last 2 paragraphs)

i think the same conclusions can and should be drawn. plotting them on the same axis makes the differences apparent way quicker, though.