T4D1 Discussion

```
Arizona Population in 2020
```

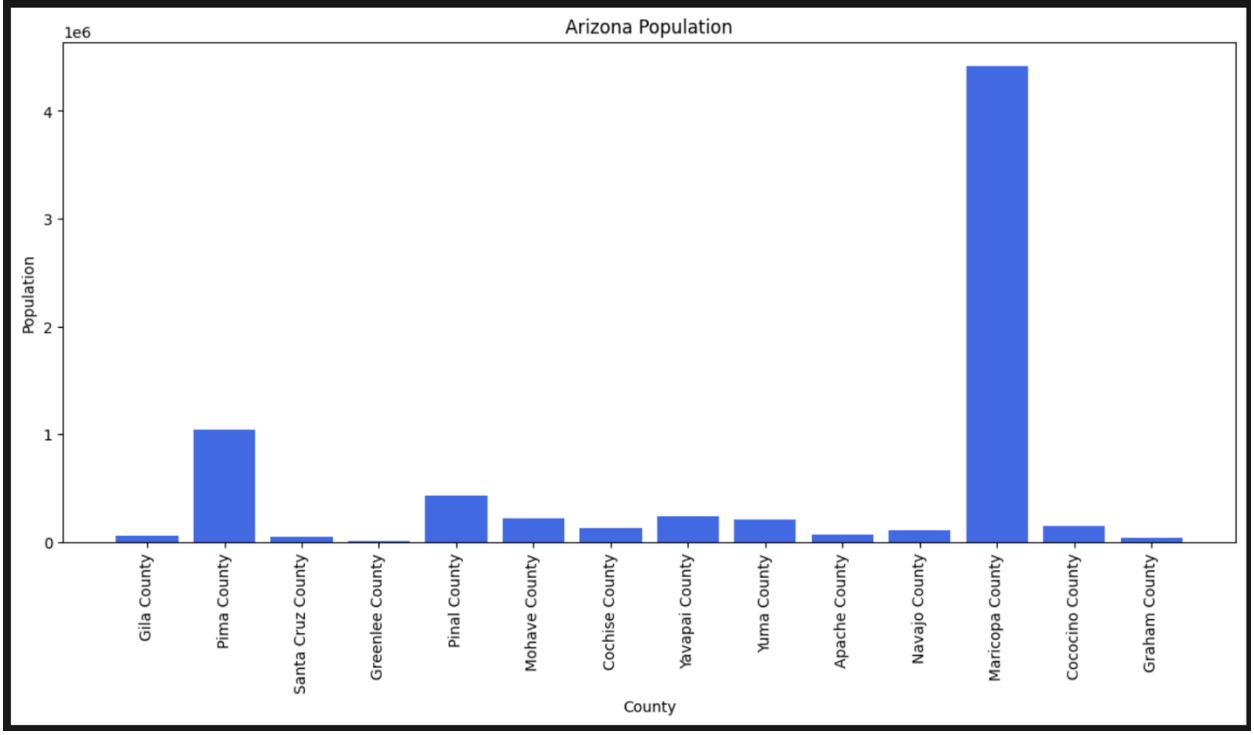
```
# Import Libraries
import pandas as pd
import matplotlib.pyplot as plt

# Load the CSV file into a DataFrame
csv_file = "AZ_Pop2020.csv"
# Read the CSV file
data = pd.read_csv(csv_file)

# Create a DataFrame for better readability
df = pd.DataFrame({
    'County' : data['Arizona Counties'],
    'Population' : data['2020'],

})
# Print the dataframe
print(df)

print('Visualisation 1')
# Create a bar graph
plt.figure(figsize=(14, 6))
# Plot and label axis
plt.bar(df['County'], df['Population'], color='royalblue')
plt.xlabel("County")
plt.ylabel("Population")
plt.title("Arizona Population")

# Rotate the x-axis labels for better readability
plt.xticks(rotation=90)

# Show the plot
plt.tight_layout
plt.show()
```

```python
print('Visualisation 2')
# Create a new df that is sorted
df2 = df.sort_values(by='Population', ascending=True)
# Plot and label table
plt.figure(figsize=(10,10))
bars = plt.barh(df2['County'], df2['Population'], color='royalblue')
plt.ylabel("Arizona Counties")
plt.xlabel('Population in 2020')
plt.title('Arizona Population 2020')
# Print out the population next to the bar
for bar, population in zip(bars, df2['Population']):
    plt.text(bar.get_width(), bar.get_y() + bar.get_height() / 2, f'{population:,}', ha='left', va='center')
# Display table
plt.tight_layout()
plt.show()
```

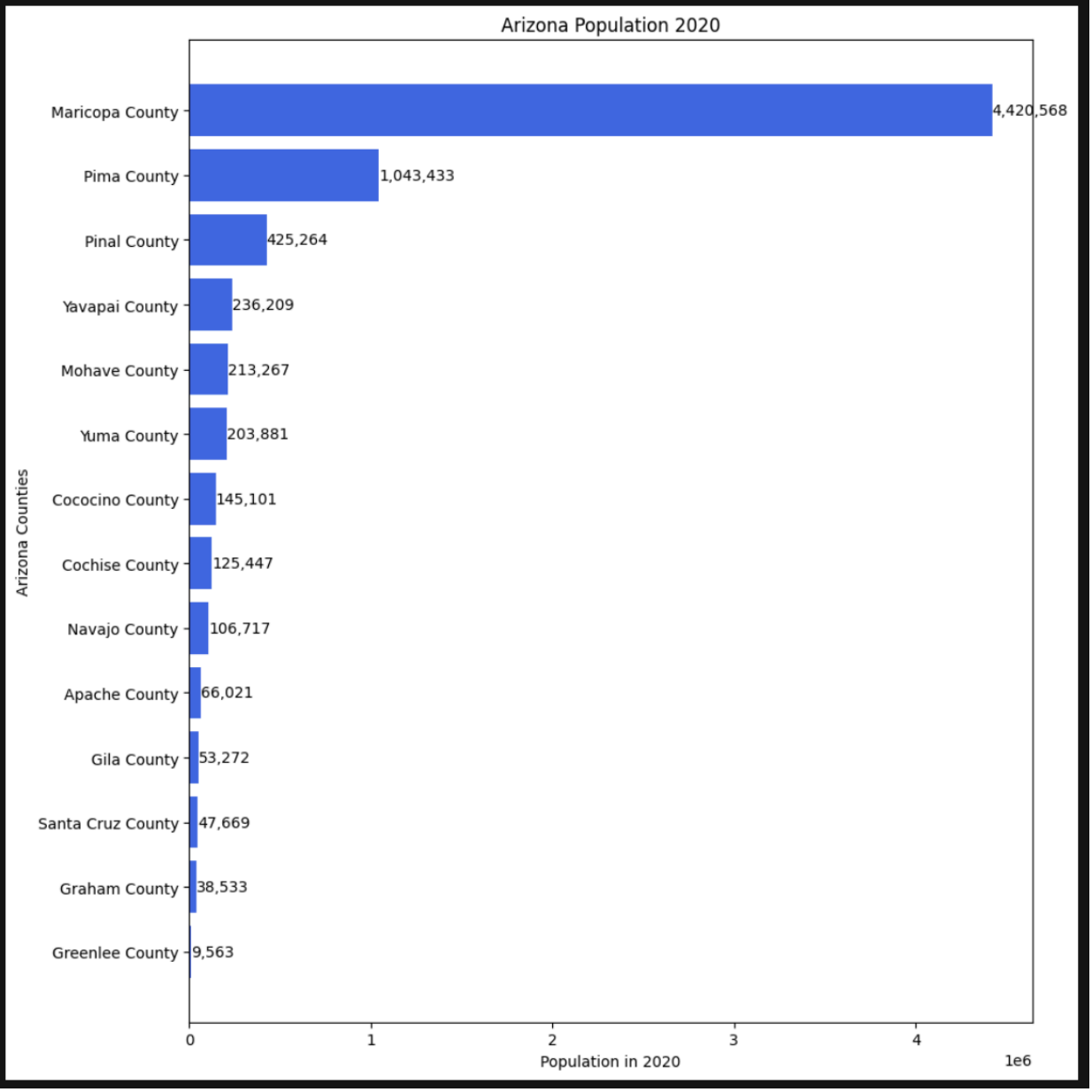
✓ 3.0s

	County	Population
0	Gila County	53272
1	Pima County	1043433
2	Santa Cruz County	47669
3	Greenlee County	9563
4	Pinal County	425264
5	Mohave County	213267
6	Cochise County	125447
7	Yavapai County	236209
8	Yuma County	203881
9	Apache County	66021
10	Navajo County	106717
11	Maricopa County	4420568
12	Cococino County	145101
13	Graham County	38533

Visualisation 1



Visualisation 2



The difference between the two data visualization shows how easily it could manipulate and mislead viewers.
In the first visualisation, labels are not properly labeled which could lead misconception that the data could be taken from last year or this year.
Another, we could state that because we could barely see the bar graph for Greenlee County, then it might be that county is inhabited.
It is also harder for viewers to understand the visualisation when it is cluttered.
In the second visualisation, we corrected these mistakes to show that this data was 2020 to lessen misinterpretation.
We added a clearer label and also added labels on the bar graph to show the population itself. It is also sorted by the population sizing in ascending order.