

R/Python Refresher:

Write a summary of your data and identify at least two questions to explore visually with your data:

COVID-19 was first identified in Wuhan, the capital of China's Hubei province. People developed common flu like symptoms. The virus has shown evidence of human-to-human transmission. Initially, WHO or Medical organizations were not able to gaze the situation. Transmission rate escalated to worrisome numbers by late January 2020 though the first case was identified in December 2019.

By mid 2020, situation around the world uncontrollable and Scientists have to work 24x7 to invent vaccines and medicines to bring situation under control. World economy was badly hit and many industries went into losses leaving many people jobless and economically hit. Situation started improving only once vaccines were fast tracked and administered to the people around the world.

I have chosen this data as this is the most worrisome issue which the world has faced in the last few decades. I wanted to show which country was badly hit.

Two questions which needs to be answered are:

1. Which country was badly hit by the pandemic?
2. Which country has done better with vaccinations?

Data sources used: Kaggle

Load Data:

```
In [1]: #Import libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

```
In [2]: #Read CSV file for covid cases, deaths and recoveries
df_covid = pd.read_csv("country_wise_latest.csv")
df_covid.head()
```

```
Out[2]:
```

| | Country/Region | Confirmed | Deaths | Recovered | Active | New cases | New deaths | New recovered | Deaths /100 Cases | Recovered /100 Cases | Deaths /100 Recovered | Confirmed last week | 1 w char |
|---|----------------|-----------|--------|-----------|--------|-----------|------------|---------------|-------------------|----------------------|-----------------------|---------------------|----------|
| 0 | Afghanistan | 36263 | 1269 | 25198 | 9796 | 106 | 10 | 18 | 3.50 | 69.49 | 5.04 | 35526 | ... |
| 1 | Albania | 4880 | 144 | 2745 | 1991 | 117 | 6 | 63 | 2.95 | 56.25 | 5.25 | 4171 | ... |
| 2 | Algeria | 27973 | 1163 | 18837 | 7973 | 616 | 8 | 749 | 4.16 | 67.34 | 6.17 | 23691 | ... |
| 3 | Andorra | 907 | 52 | 803 | 52 | 10 | 0 | 0 | 5.73 | 88.53 | 6.48 | 884 | ... |
| 4 | Angola | 950 | 41 | 242 | 667 | 18 | 1 | 0 | 4.32 | 25.47 | 16.94 | 749 | ... |

```
In [3]: #Read CSV file for covid vaccinations
df_vaccines = pd.read_csv("vaccinations.csv")
df_vaccines.drop_duplicates(subset='location',keep='first',inplace=True)
df_vaccines.head()
```

```
Out[3]:
```

| | location | iso_code | date | total_vaccinations | people_vaccinated | people_fully_vaccinated | total_boosters | daily_vaccinations |
|------|-------------|----------|------------|--------------------|-------------------|-------------------------|----------------|--------------------|
| 0 | Afghanistan | AFG | 2021-02-22 | 0.0 | 0.0 | NaN | NaN | NaN |
| 575 | Africa | OWID_AFR | 2021-01-09 | 0.0 | 0.0 | NaN | NaN | NaN |
| 1199 | Albania | ALB | 2021-01-10 | 0.0 | 0.0 | NaN | NaN | NaN |
| 1802 | Algeria | DZA | 2021-01-29 | 0.0 | 0.0 | NaN | NaN | NaN |
| 2386 | Andorra | AND | 2021-01-25 | 576.0 | 576.0 | NaN | NaN | NaN |

```
In [4]: #Sorting the dataframe
df_vaccines_sorted= df_vaccines.sort_values(['total_vaccinations'])
df_vaccines.head()
```

```
Out[4]:
```

| | location | iso_code | date | total_vaccinations | people_vaccinated | people_fully_vaccinated | total_boosters | daily_vaccinations |
|------|-------------|----------|------------|--------------------|-------------------|-------------------------|----------------|--------------------|
| 0 | Afghanistan | AFG | 2021-02-22 | 0.0 | 0.0 | NaN | NaN | NaN |
| 575 | Africa | OWID_AFR | 2021-01-09 | 0.0 | 0.0 | NaN | NaN | NaN |
| 1199 | Albania | ALB | 2021-01-10 | 0.0 | 0.0 | NaN | NaN | NaN |
| 1802 | Algeria | DZA | 2021-01-29 | 0.0 | 0.0 | NaN | NaN | NaN |
| 2386 | Andorra | AND | 2021-01-25 | 576.0 | 576.0 | NaN | NaN | NaN |

```
In [5]: #Sorting the dataframe
df_covid_sorted= df_covid.sort_values(['Confirmed','Recovered','Deaths'])
df_covid_sorted.head()
```

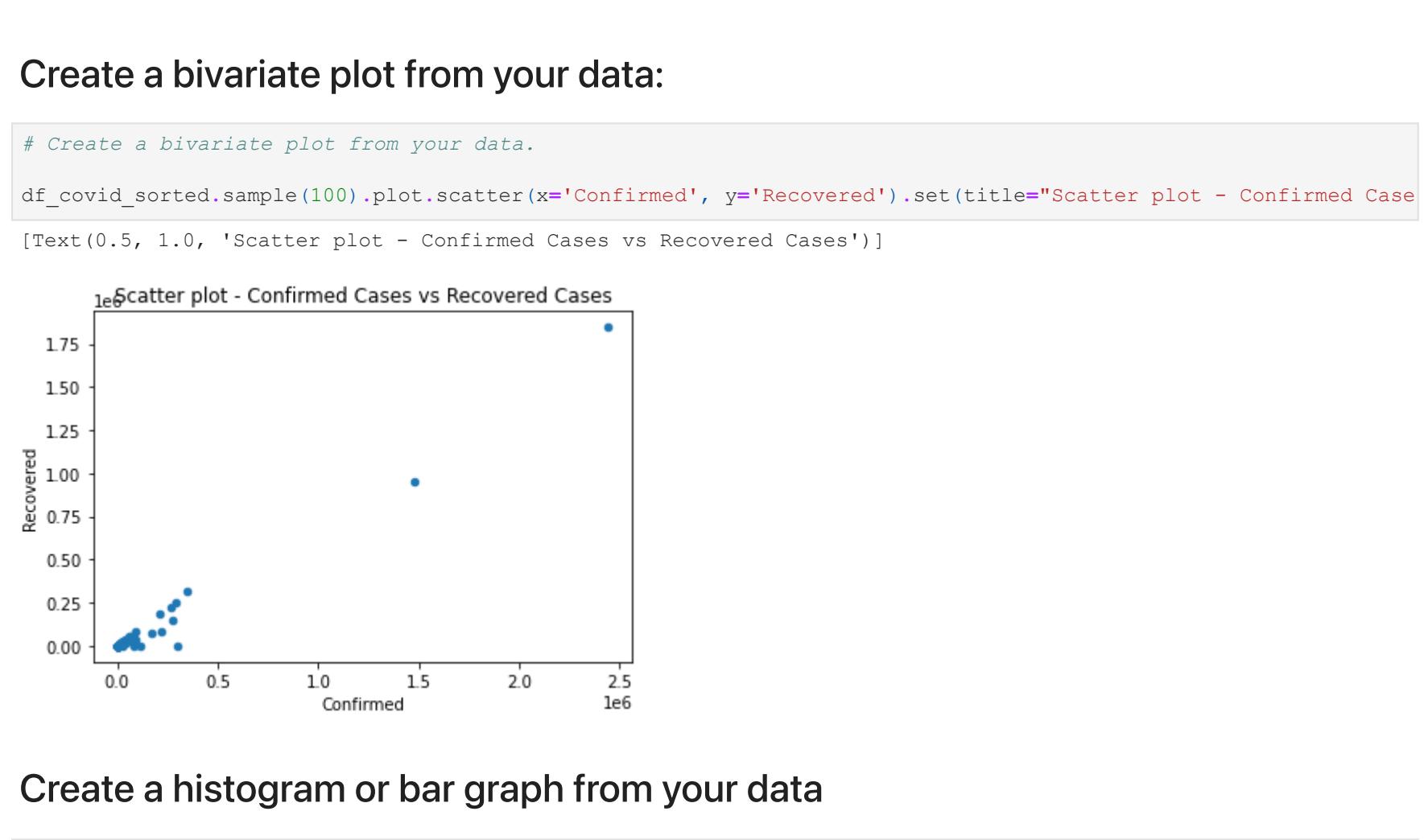
```
Out[5]:
```

| | Country/Region | Confirmed | Deaths | Recovered | Active | New cases | New deaths | New recovered | Deaths /100 Cases | Recovered /100 Cases | Deaths /100 Recovered | Confirmed last week | 1 w char |
|-----|-----------------------|-----------|--------|-----------|--------|-----------|------------|---------------|-------------------|----------------------|-----------------------|---------------------|----------|
| 183 | Western Sahara | 10 | 1 | 8 | 1 | 0 | 0 | 0 | 10.0 | 80.00 | 12.5 | 10 | ... |
| 75 | Holy See | 12 | 0 | 12 | 0 | 0 | 0 | 0 | 0.0 | 100.00 | 0.0 | 12 | ... |
| 68 | Greenland | 14 | 0 | 13 | 1 | 1 | 0 | 0 | 0.0 | 92.86 | 0.0 | 13 | ... |
| 140 | Saint Kitts and Nevis | 17 | 0 | 15 | 2 | 0 | 0 | 0 | 0.0 | 88.24 | 0.0 | 17 | ... |
| 49 | Dominica | 18 | 0 | 18 | 0 | 0 | 0 | 0 | 0.0 | 100.00 | 0.0 | 18 | ... |

Create a histogram or bar graph from your data.

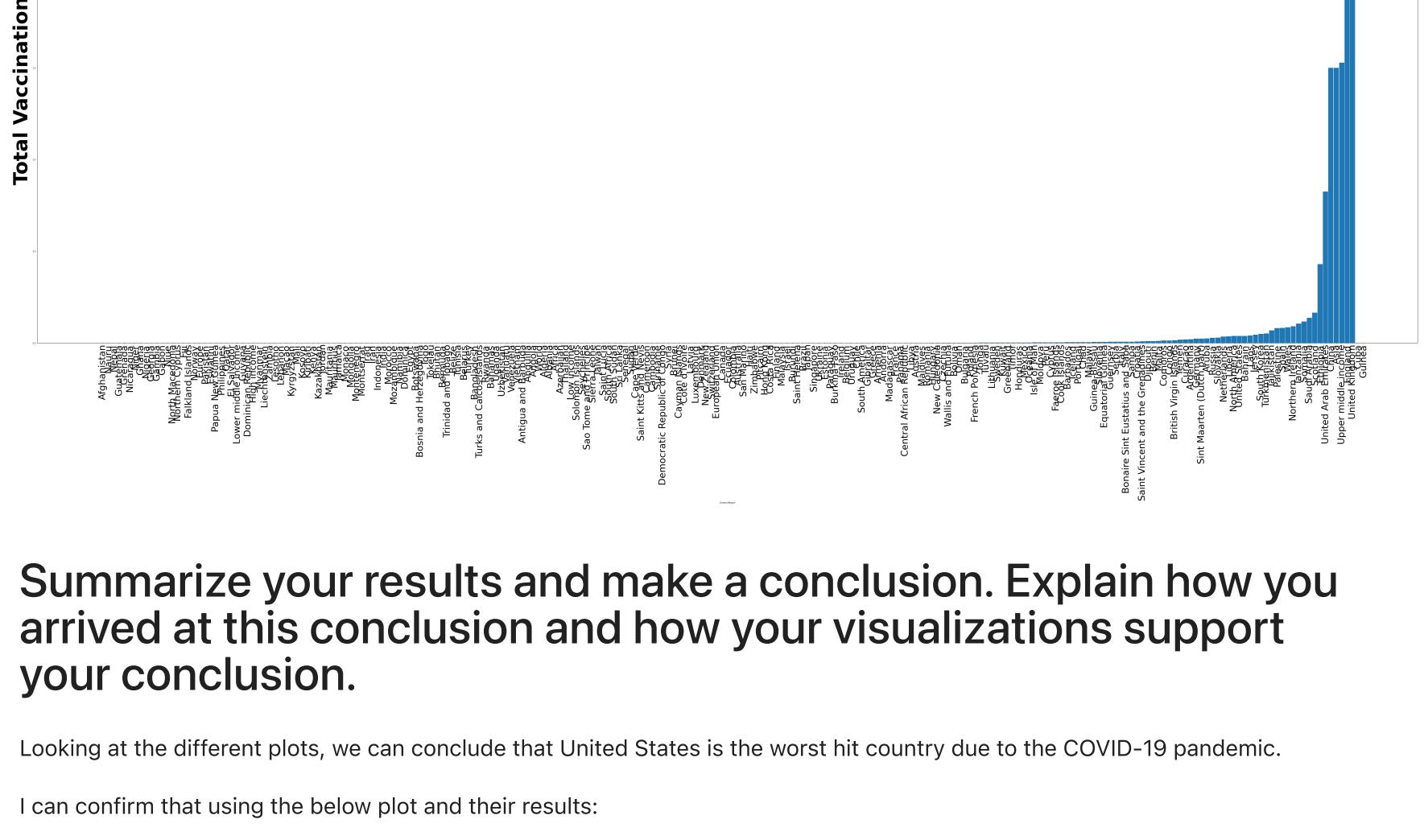
```
In [6]: plt.figure(figsize=(120,46))
# bar plot with matplotlib
width = 0.8 # the width of the bars
plt.bar('Country/Region', 'Confirmed', width,data=df_covid_sorted)
plt.xlabel("Country/Region")
plt.xticks(rotation=90, horizontalalignment="center", size=48)
plt.ylabel("Confirmed COVID cases", size=95, weight='bold')
plt.title("Bar plot in Ascending Order - Confirmed Cases per Country", size=98, weight='bold')
plt.show()
```

Bar plot in Ascending Order - Confirmed Cases per Country



```
In [7]: plt.figure(figsize=(120,46))
# bar plot with matplotlib
width = 0.8 # the width of the bars
plt.bar('Country/Region', 'Deaths', width,data=df_covid_sorted)
plt.xlabel("Country/Region")
plt.xticks(rotation=90, horizontalalignment="center", size=48)
plt.ylabel("COVID Deaths", size=95, weight='bold')
plt.title("Bar plot in Ascending Order - COVID Deaths per Country", size=98, weight='bold')
plt.show()
```

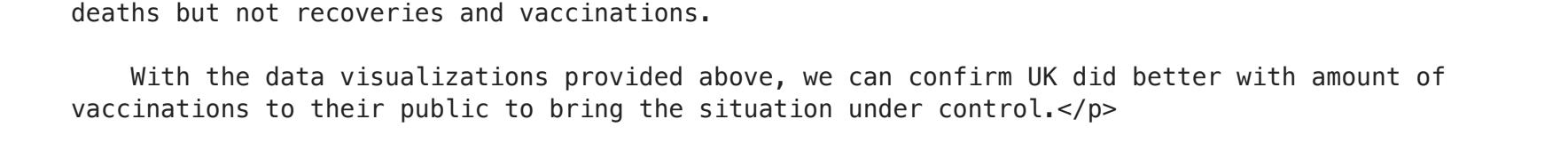
Bar plot in Ascending Order - COVID Deaths per Country



Create a boxplot from your data:

```
In [8]: #Create a boxplot from your data.
df_covid_filtered = df_covid_sorted[['Country/Region','Confirmed','Deaths','Recovered','Active']]
df_covid_filtered.plot(kind='box', figsize = (10,6), subplots = True)
```

Box plot from your data



Create a bivariate plot from your data:

```
In [10]: # Create a bivariate plot from your data.
df_covid_sorted.sample(100).plot.scatter(x='Confirmed', y='Recovered').set(title="Scatter plot - Confirmed Cases vs Recovered Cases")
```


Create a histogram or bar graph from your data

```
In [11]: plt.figure(figsize=(120,46))
# bar plot with matplotlib
width = 0.95 # the width of the bars
plt.bar('location', 'total_vaccinations',width,data=df_vaccines_sorted)
plt.xlabel("Country/Region")
plt.xticks(rotation=90, horizontalalignment="center", size=45)
plt.ylabel("Total Vaccinations", size=95, weight='bold')
plt.title("Bar plot in Ascending Order - Total Vaccinations per Country", size=98, weight='bold')
plt.show()
```

Bar plot in Ascending Order - Total Vaccinations per Country

Summarize your results and make a conclusion. Explain how you arrived at this conclusion and how your visualizations support your conclusion.

Looking at the different plots, we can conclude that United States is the worst hit country due to the COVID-19 pandemic.

I can confirm that using the below plot and their results:

1. Bar plot in Ascending Order - Confirmed Cases per Country: This plot confirms United States had the most number of confirmed COVID cases.
2. Bar plot in Ascending Order - COVID Deaths per Country: This plot confirms United States had the most number of COVID deaths.
3. Bar plot in Ascending Order - Recovered Cases per Country: This plot confirms Brazil had the most number of COVID recoveries compared to United States but United States has the most number confirmed cases and deaths.
4. Bar plot in Ascending Order - Total Vaccinations per Country: This plot confirms UK had the most number of COVID vaccinations compared to United States but United States has the most number confirmed cases and deaths.
5. Boxplot from your data: Box plot shows that the recoveries are much more than the deaths globally, so we can confirm, overall recoveries globally are good.
6. Bivariate plot from your data: Scatter plot is one of the bivariate plots. Scatter plot shows that Confirmed cases vs Recoveries is linear which means globally situation was not best or worst.

Conclusion:

Though the situation around was neither good nor bad. However, looking at all the plots, we can confirm United States is the worst affected country which has more confirmed cases and deaths but not recoveries and vaccinations.

With the data visualizations provided above, we can confirm UK did better with amount of vaccinations to their public to bring the situation under control.

