

Assignment 4 – Colormap

Weilin Ruan 50018083 Group5

1. Data Attribute Type for 'new_cases':

- **Type:** Quantitative (Numerical)
- **Explanation:** The new_cases attribute represents the number of new COVID-19 cases reported on a given day. Since it is a numerical value that can vary continuously, it is considered a quantitative attribute.

2. Colormap Type for 'new_cases':

- **Type:** Sequential
- **Explanation:** A sequential colormap is appropriate for encoding quantitative data where the values have a natural order. For new_cases, a sequential colormap can effectively represent the magnitude of new cases, with lighter colors for lower values and darker colors for higher values.

3. Data Attribute Type for 'location':

- **Type:** Categorical (Nominal)
- **Explanation:** The location attribute represents different countries or regions. Since these locations do not have a natural order, they are considered categorical attributes.

4. Colormap Type for 'location':

- **Type:** Categorical
- **Explanation:** A categorical colormap is appropriate for encoding categorical data where the values do not have a natural order. For location, a categorical colormap can effectively represent different countries or regions with distinct colors.

Coding

To create the scatter plot, we will use Python with libraries such as pandas for data manipulation and matplotlib or seaborn for visualization.

```
1. import pandas as pd
2. import plotly.express as px
3. import plotly.io as pio
4. pio.renderers.default = 'browser'
5.
6. # Load the data
```

```

7. #url = "https://github.com/owid/covid-19-
   data/tree/master/public/data/jhu/new_cases.csv"
8. #data = pd.read_csv(url)
9. df = pd.read_csv("full_data-2.csv")
10.
11. # Get the latest date in the data
12. latest_date = df['date'].max()
13.
14. # Filter the data for the latest date
15. latest_data = df[df['date'] == latest_date].copy()
16.
17. # Fill NaN values in the 'new_cases' column with 0 or any a
   ppropriate value
18. latest_data['new_cases'] = latest_data['new_cases'].fillna(
   0)
19.
20. # Calculate case-
   fatality rate (optional, just for additional info)
21. latest_data['case_fatality_rate'] = (latest_data['total_dea
   ths'] / latest_data['total_cases']) * 100
22.
23. print(df.head())
24. #print(latest_data)
25.
26. # Create a scatter plot for total_cases vs total_deaths
27. # Assign colors based on new_cases using a sequential color
   map
28. fig = px.scatter(
29.     latest_data,
30.     x='total_cases',
31.     y='total_deaths',
32.     text='location',
33.     color='new_cases', # Sequential colormap based on 'new
   _cases'
34.     size='new_cases', # Adjust point size based on 'new_ca
   ses'
35.     hover_data={
36.         'location': True,
37.         'case_fatality_rate': '::.2f',
38.         'total_cases': True,
39.         'total_deaths': True,
40.         'new_cases': True
41.     },

```

```

42.     labels={'total_cases': 'Total Cases', 'total_deaths': '
      Total Deaths'},
43.     title='COVID-
      19 Total Cases vs Total Deaths (Colored by New Cases)',
44.     log_x=True,
45.     log_y=True,
46.     color_continuous_scale='Viridis' # Sequential colormap
      for new_cases
47.)
48.
49.# Adjust the layout for better display
50.fig.update_traces(marker=dict(opacity=0.7))
51.fig.update_layout(
52.     hovermode='closest',
53.     width=2400, # Set the width of the figure
54.     height=1600, # Set the height of the figure
55.)
56.
57.# Show the plot
58.fig.show()

```

Results:

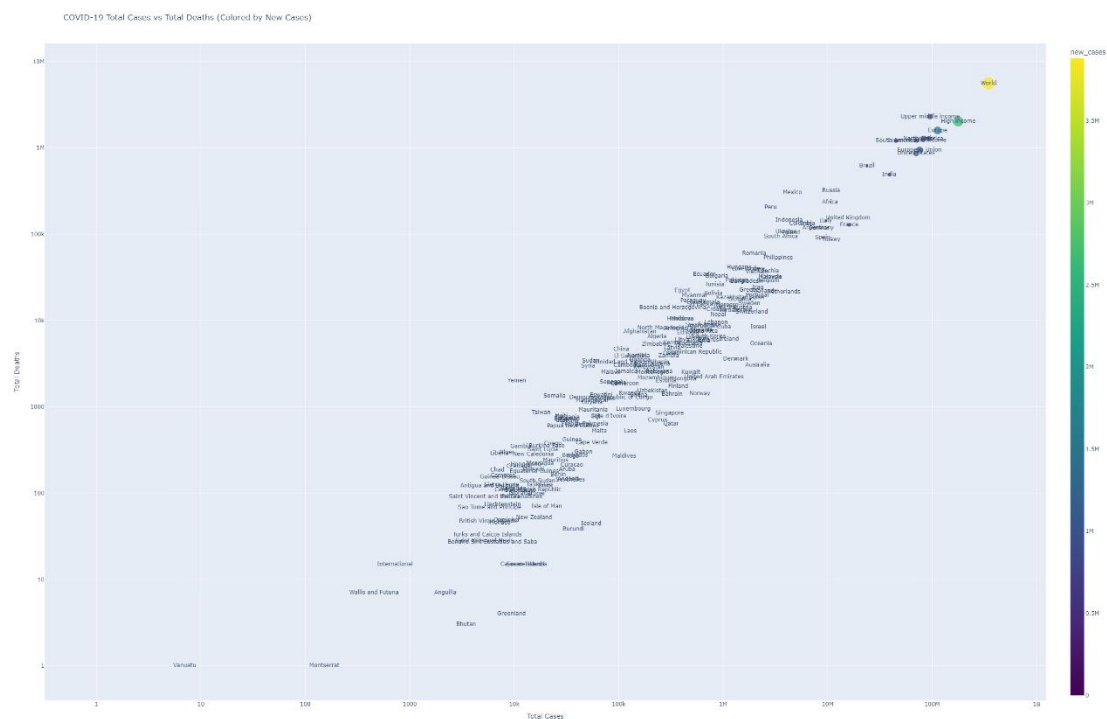


Fig 1 Overview of the Colormap

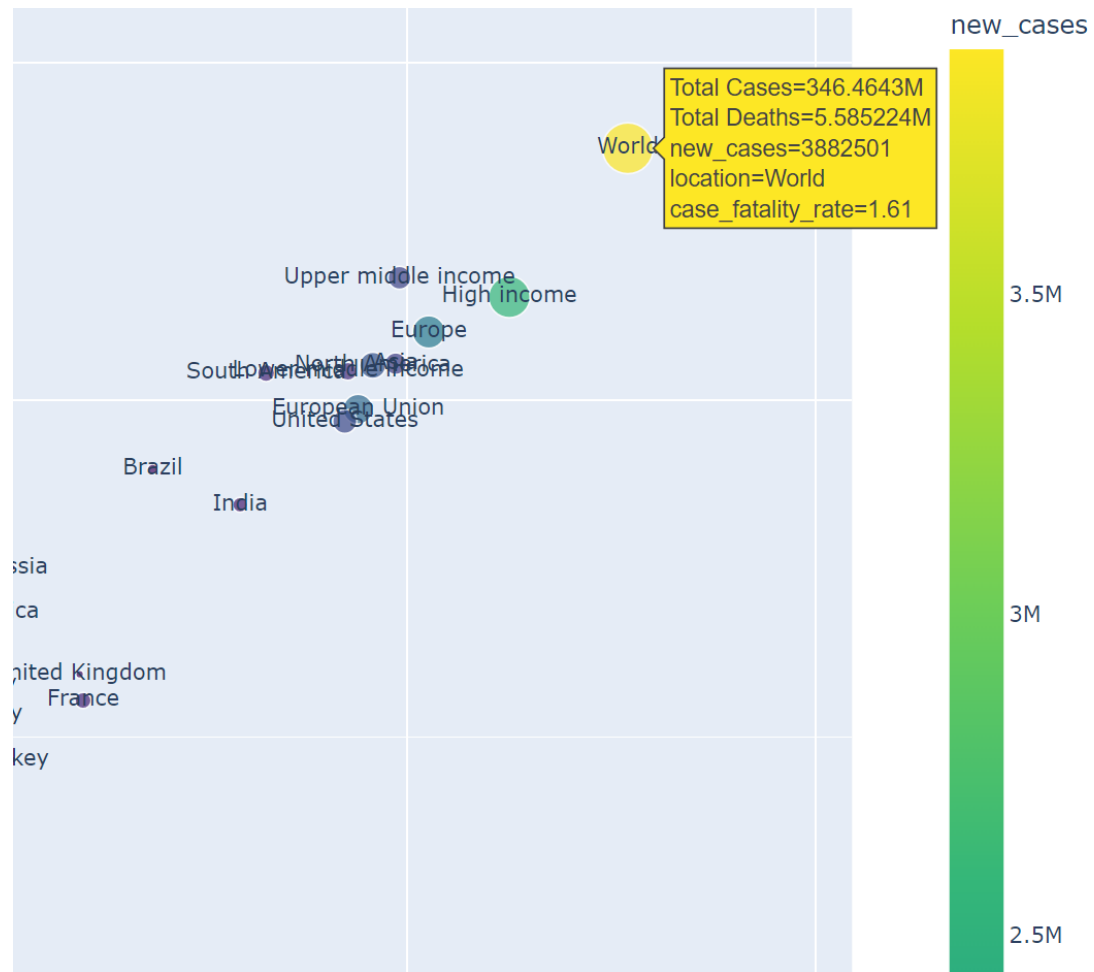


Figure 1 Interaction of the Colormap

Explanation of the Code

1. Import Libraries: Import necessary libraries for data manipulation and visualization.
2. Load Data: Load the data from the provided URL.
3. Filter Latest Data: Filter the data to get the latest date's records.
4. Create Scatter Plot: Initialize a figure for the scatter plot.
5. Assign Colors Based on new_cases:
 - Use a sequential colormap (Reds) to represent the magnitude of new cases.
 - Normalize the new_cases values to map them to the colormap.
 - Create the scatter plot with colors based on new_cases.
6. Assign Colors Based on location:
 - Use a categorical colormap (tab20) to represent different locations.

- Map each location to a unique color.
 - Create the scatter plot with colors based on location.
7. Adjust Point Size Based on new_cases:
- Scale the point size by dividing new_cases by 100 to make the points visible but not too large.
8. Add Colorbar and Legend:
- Add a colorbar to show the mapping of colors to new_cases.
 - Add a legend to show the mapping of colors to locations.
9. Set Labels and Title: Set the x-axis, y-axis labels, and the title of the plot.
10. Show the Plot: Display the scatter plot.

This code will generate a scatter plot where the x-axis represents total cases, the y-axis represents total deaths, the colors represent either new cases or locations, and the point sizes represent the magnitude of new cases.