Final Project

Dylan Xia Kevin Yao Yuqing Wen

2024-12-01

- Partner 1 : Dylan Xia; dizhexia
- Partner 2 : Kevin Yao; qiyin
- Partner 3 : Yuqing Wen, wyuqing

```
import pandas as pd
import geopandas as gpd
import matplotlib.pyplot as plt
from shapely.geometry import shape
import requests
import altair as alt

import warnings
warnings.filterwarnings('ignore')
alt.renderers.enable("png")
```

```
RendererRegistry.enable('png')
```

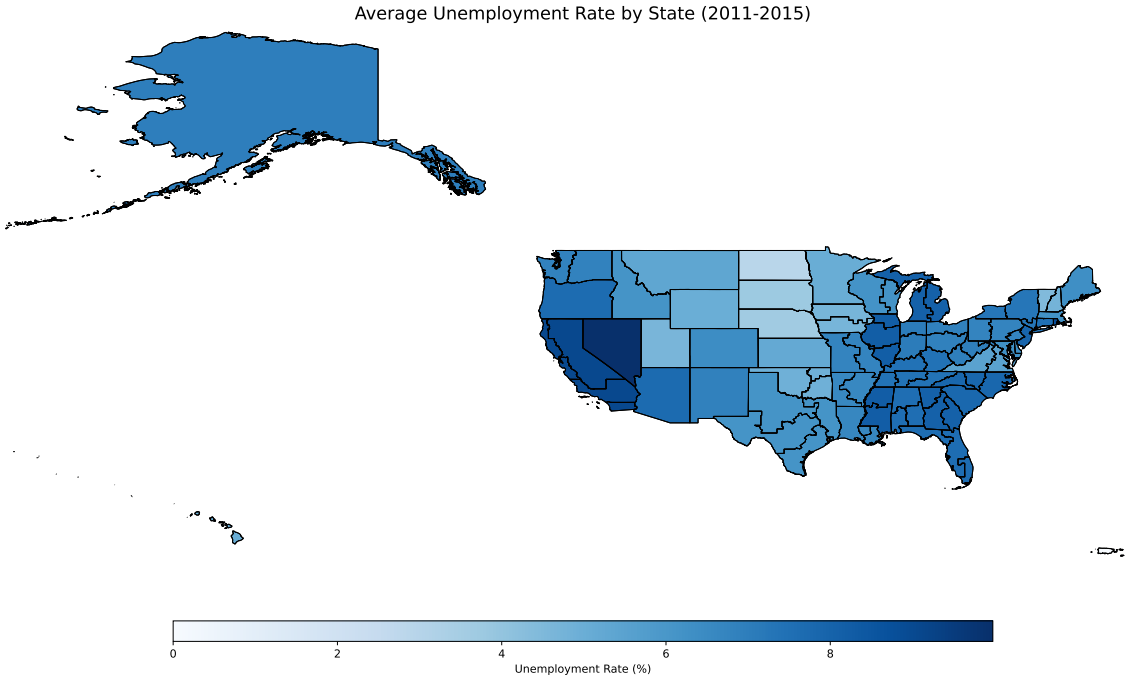
Step 1: Plot data based on Unemployment Rate

Unemployment Rate (2011~2015), Map by State

```
# Step 1: Load the unemployment data
data_path = "state_yearly_unemployment_rate_with_state_name.csv"
unemployment_data = pd.read_csv(data_path)

# Step 2: Calculate the average unemployment rate (2011-2015) by state
# Assuming the dataset contains columns: 'state_name', 'year',
 ↪  'unemployment_rate'
```

```
unemployment_avg = (
    unemployment_data[(unemployment_data['year'] >= 2011) &
 ↪  (unemployment_data['year'] <= 2015)]
    .groupby('state_name')['unemployment_rate']
    .mean()
    .reset_index()
)
# Convert unemployment rate to percentage
unemployment_avg['unemployment_rate'] = unemployment_avg['unemployment_rate']
 ↪  * 100

# Step 3: Load the JSON data from the API endpoint for state boundaries
shape_url = 'https://data.ojp.usdoj.gov/resource/5fdt-n5ne.json'
response = requests.get(shape_url)

if response.status_code == 200:
    geo_data = response.json()
else:
    raise ValueError(f"Failed to fetch GeoJSON data. HTTP Status Code:
     ↪  {response.status_code}")

# Step 4: Convert JSON data to a GeoDataFrame
geometries = [shape(feature["the_geom"]) for feature in geo_data]
properties = [{key: feature[key] for key in feature if key != "the_geom"} for
 ↪  feature in geo_data]
shape_data = gpd.GeoDataFrame(properties, geometry=geometries)

# Step 5: Merge shape data with unemployment rate data
unemployment_avg['state_name'] = unemployment_avg['state_name'].str.strip()
merged_data = shape_data.merge(unemployment_avg, left_on='state',
 ↪  right_on='state_name', how='left')

# Fill missing values for states with no data
merged_data['unemployment_rate'] = merged_data['unemployment_rate'].fillna(0)

# Step 6: Plot the unemployment map
fig, ax = plt.subplots(1, 1, figsize=(15, 10))

# Plot unemployment data on the map
merged_data.plot(
    column='unemployment_rate',
    cmap='Blues',
```

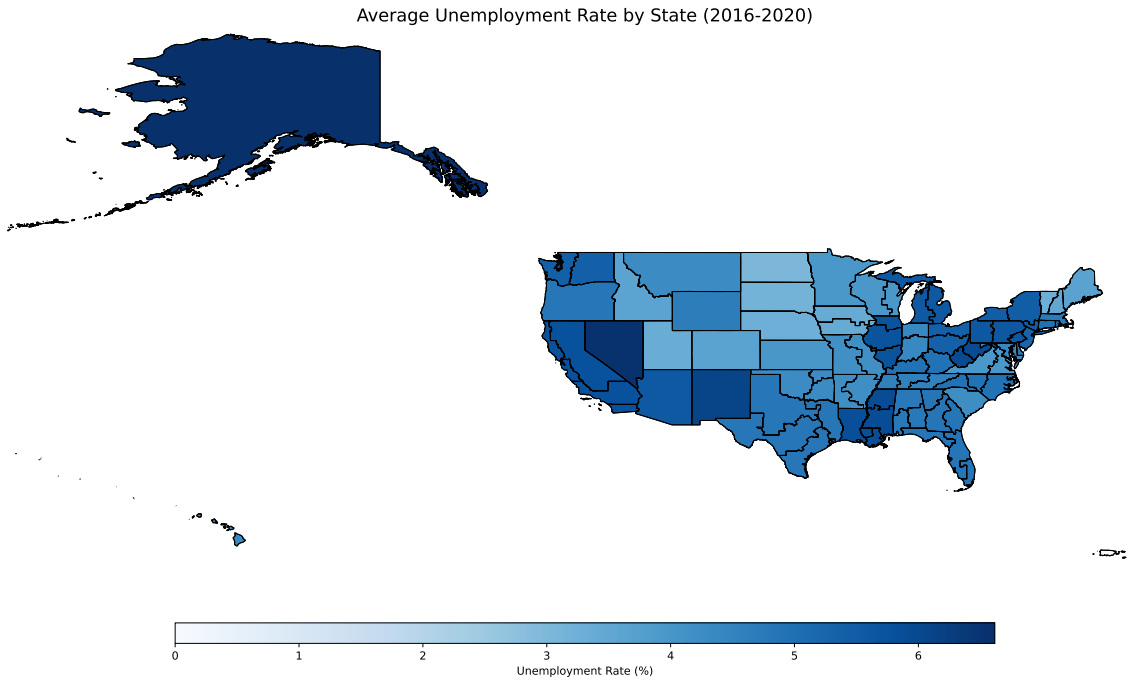
```
    linewidth=0.8,
    ax=ax,
    edgecolor='black',
    legend=True,
    legend_kwds={
        'shrink': 0.7,
        'orientation': "horizontal",
        'pad': 0.05,
        'aspect': 40,
        'label': "Unemployment Rate (%)"
    }
)

# Add state boundaries
merged_data.boundary.plot(ax=ax, linewidth=0.8, color="black")

# Adjust map extent to include Alaska and Hawaii
ax.set_xlim([-180, -60])  # Extends to cover Alaska and Hawaii
ax.set_ylim([15, 72])     # Adjusted for both Hawaii and Alaska latitudes

# Customize the plot further for better visualization
ax.set_title('Average Unemployment Rate by State (2011-2015)', fontsize=16)
ax.set_axis_off()
plt.tight_layout()

plt.show()
```



Unemployment Rate (2016~2020), Map by State

```
# Calculate the average unemployment rate (2016-2020) by state
# Assuming the dataset contains columns: 'state_name', 'year',
 ↪  'unemployment_rate'
unemployment_avg = (
    unemployment_data[(unemployment_data['year'] >= 2016) &
 ↪  (unemployment_data['year'] <= 2020)]
    .groupby('state_name')['unemployment_rate']
    .mean()
    .reset_index()
)
# Convert unemployment rate to percentage
unemployment_avg['unemployment_rate'] = unemployment_avg['unemployment_rate']
 ↪  * 100

# Merge shape data with unemployment rate data
unemployment_avg['state_name'] = unemployment_avg['state_name'].str.strip()
merged_data = shape_data.merge(unemployment_avg, left_on='state',
 ↪  right_on='state_name', how='left')
```

```
# Fill missing values for states with no data
merged_data['unemployment_rate'] = merged_data['unemployment_rate'].fillna(0)

# Plot the unemployment map
fig, ax = plt.subplots(1, 1, figsize=(15, 10))

# Plot unemployment data on the map
merged_data.plot(
    column='unemployment_rate',
    cmap='Blues',
    linewidth=0.8,
    ax=ax,
    edgecolor='black',
    legend=True,
    legend_kwds={
        'shrink': 0.7,
        'orientation': "horizontal",
        'pad': 0.05,
        'aspect': 40,
        'label': "Unemployment Rate (%)"
    }
)

# Add state boundaries
merged_data.boundary.plot(ax=ax, linewidth=0.8, color="black")

# Adjust map extent to include Alaska and Hawaii
ax.set_xlim([-180, -60])  # Extends to cover Alaska and Hawaii
ax.set_ylim([15, 72])     # Adjusted for both Hawaii and Alaska latitudes

# Customize the plot further for better visualization
ax.set_title('Average Unemployment Rate by State (2016-2020)', fontsize=16)
ax.set_axis_off()
plt.tight_layout()

plt.show()
```



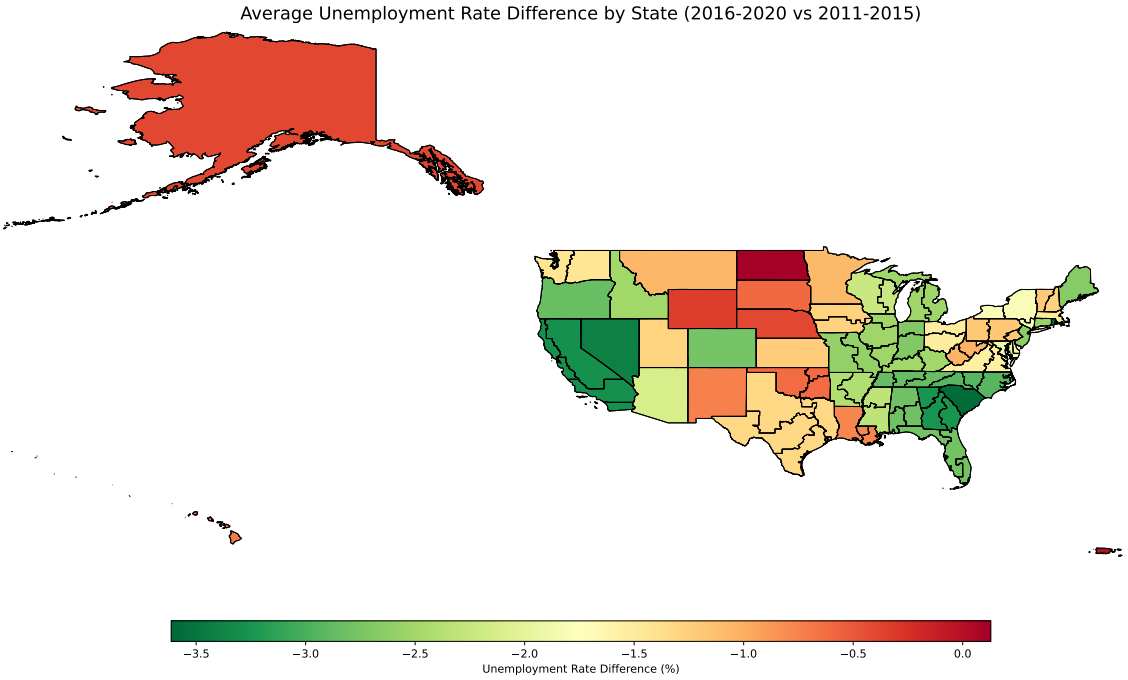
Unemployment Rate (Difference), Map by State

```
import pandas as pd
import geopandas as gpd
import matplotlib.pyplot as plt
from shapely.geometry import shape
import requests

# Step 1: Load the unemployment data
data_path = "state_yearly_unemployment_rate_with_state_name.csv"
unemployment_data = pd.read_csv(data_path)

# Step 2: Calculate average unemployment rates for 2011-2015 and 2016-2020 by
 ↪  state
unemployment_avg_2011_2015 = (
    unemployment_data[(unemployment_data['year'] >= 2011) &
 ↪  (unemployment_data['year'] <= 2015)]
    .groupby('state_name')['unemployment_rate']
    .mean()
    .reset_index()
```

```
)
unemployment_avg_2016_2020 = (
    unemployment_data[(unemployment_data['year'] >= 2016) &
 ↪  (unemployment_data['year'] <= 2020)]
    .groupby('state_name')['unemployment_rate']
    .mean()
    .reset_index()
)

# Convert to percentage
unemployment_avg_2011_2015['unemployment_rate'] =
 ↪  unemployment_avg_2011_2015['unemployment_rate'] * 100
unemployment_avg_2016_2020['unemployment_rate'] =
 ↪  unemployment_avg_2016_2020['unemployment_rate'] * 100

# Step 3: Calculate the difference between 2016-2020 and 2011-2015
unemployment_diff = pd.merge(
    unemployment_avg_2016_2020,
    unemployment_avg_2011_2015,
    on='state_name',
    suffixes=('_2016_2020', '_2011_2015')
)
unemployment_diff['rate_difference'] = (
    unemployment_diff['unemployment_rate_2016_2020'] -
 ↪  unemployment_diff['unemployment_rate_2011_2015']
)

# Step 4: Load the JSON data from the API endpoint for state boundaries
shape_url = 'https://data.ojp.usdoj.gov/resource/5fdt-n5ne.json'
response = requests.get(shape_url)

if response.status_code == 200:
    geo_data = response.json()
else:
    raise ValueError(f"Failed to fetch GeoJSON data. HTTP Status Code:
     ↪  {response.status_code}")

# Step 5: Convert JSON data to a GeoDataFrame
geometries = [shape(feature["the_geom"]) for feature in geo_data]
properties = [{key: feature[key] for key in feature if key != "the_geom"} for
 ↪  feature in geo_data]
shape_data = gpd.GeoDataFrame(properties, geometry=geometries)
```

```
# Step 6: Merge shape data with unemployment difference data
unemployment_diff['state_name'] = unemployment_diff['state_name'].str.strip()
merged_data = shape_data.merge(unemployment_diff, left_on='state',
 ↪  right_on='state_name', how='left')

# Fill missing values for states with no data
merged_data['rate_difference'] = merged_data['rate_difference'].fillna(0)

# Step 7: Plot the unemployment difference map
fig, ax = plt.subplots(1, 1, figsize=(15, 10))

# Plot unemployment difference using a diverging colormap
merged_data.plot(
    column='rate_difference',
    cmap='RdYlGn_r',  # Red-White-Green reversed color map
    linewidth=0.8,
    ax=ax,
    edgecolor='black',
    legend=True,
    legend_kwds={
        'shrink': 0.7,
        'orientation': "horizontal",
        'pad': 0.05,
        'aspect': 40,
        'label': "Unemployment Rate Difference (%)"
    }
)

# Add state boundaries
merged_data.boundary.plot(ax=ax, linewidth=0.8, color="black")

# Adjust map extent to include Alaska and Hawaii
ax.set_xlim([-180, -60])  # Extends to cover Alaska and Hawaii
ax.set_ylim([15, 72])     # Adjusted for both Hawaii and Alaska latitudes

# Customize the plot further for better visualization
ax.set_title('Average Unemployment Rate Difference by State (2016-2020 vs
 ↪  2011-2015)', fontsize=16)
ax.set_axis_off()
plt.tight_layout()
```

```
plt.show()
```


Average Unemployment Rate Difference by State (2016-2020 vs 2011-2015)
−3.5
−3.0
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
Unemployment Rate Difference (%)