Exercise2

March 20, 2024

[1]:
```
import geopandas as gpd
```

[2]:
```
import matplotlib.pyplot as plt
```

[3]:
```
import pandas as pd
```

[4]:
```
import mapclassify
```

0.0.1 1. Using the file ~/data/scag_region.parquet create a new GeoDataFrame that contains only the census tract data for San Diego County. Be sure to reset the index of the newly created GeoDataFrame.

[5]:
```
scag_region = gpd.read_parquet("~/data/scag_region.parquet")
```

[7]:
```
print(scag_region.columns)
```

```
Index(['geoid', 'n_asian_under_15', 'n_black_under_15', 'n_hispanic_under_15',
       'n_native_under_15', 'n_white_under_15', 'n_persons_under_18',
       'n_asian_over_60', 'n_black_over_60', 'n_hispanic_over_60',
       …
       'year', 'n_total_housing_units_sample', 'p_nonhisp_white_persons',
       'p_white_over_60', 'p_black_over_60', 'p_hispanic_over_60',
       'p_native_over_60', 'p_asian_over_60', 'p_disabled', 'geometry'],
      dtype='object', length=194)
```

[8]:
```
gdf = gpd.read_parquet("~/data/scag_region.parquet")
```

[10]:
```
gdf.shape
```

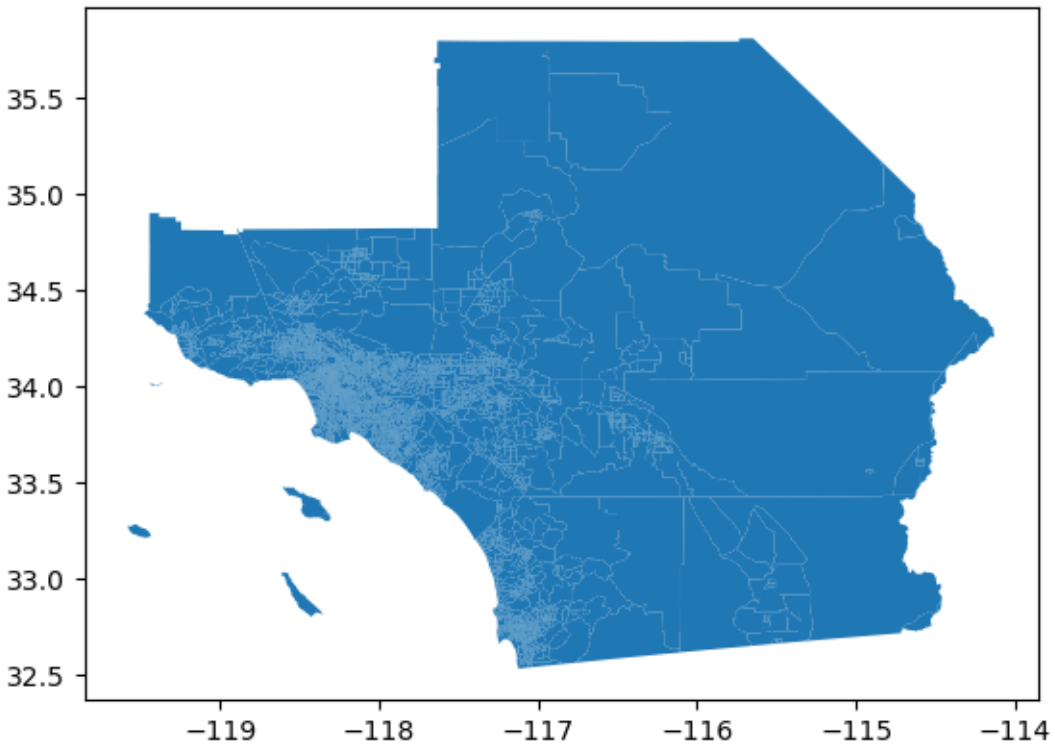
[10]:
```
(4580, 194)
```

[11]:
```
type(gdf)
```

[11]:
```
geopandas.geodataframe.GeoDataFrame
```

[13]:
```
gdf.plot()
```

[13]: <Axes: >



[14]:
```
scag_region['06073'] = scag_region['geoid'].str.slice(2, 5)
```

[15]:
```
san_diego_tracts = scag_region[scag_region['06073'] == '073'].
↪reset_index(drop=True)
```

[16]:
```
san_diego_tracts.head()
```

[16]:

	geoid	n_asian_under_15	n_black_under_15	n_hispanic_under_15	\
0	06073014901	59.0	27.0	109.0	
1	06073000300	0.0	0.0	46.0	
2	06073000800	16.0	77.0	78.0	
3	06073002201	62.0	139.0	868.0	
4	06073018509	70.0	56.0	1032.0	

	n_native_under_15	n_white_under_15	n_persons_under_18	n_asian_over_60	\
0	0.0	318.0	705.0	NaN	
1	0.0	40.0	191.0	NaN	
2	0.0	59.0	298.0	NaN	
3	0.0	0.0	1302.0	NaN	
4	0.0	83.0	1671.0	NaN	

```
   n_black_over_60  n_hispanic_over_60  …  n_total_housing_units_sample  \
0              NaN                 NaN  …                        1876.0
1              NaN                 NaN  …                        3046.0
2              NaN                 NaN  …                        2702.0
3              NaN                 NaN  …                        1321.0
4              NaN                 NaN  …                        1701.0

   p_nonhisp_white_persons  p_white_over_60  p_black_over_60  \
0                65.158807              NaN              NaN
1                72.801901              NaN              NaN
2                59.056509              NaN              NaN
3                 9.626473              NaN              NaN
4                23.131806              NaN              NaN

   p_hispanic_over_60  p_native_over_60  p_asian_over_60  p_disabled  \
0                 NaN               NaN              NaN         NaN
1                 NaN               NaN              NaN         NaN
2                 NaN               NaN              NaN         NaN
3                 NaN               NaN              NaN         NaN
4                 NaN               NaN              NaN         NaN

                                            geometry  06073
0  POLYGON ((-117.01957 32.76373, -117.01562 32.7…    073
1  POLYGON ((-117.16864 32.74897, -117.16602 32.7…    073
2  POLYGON ((-117.14632 32.74842, -117.14250 32.7…    073
3  POLYGON ((-117.11577 32.75522, -117.11362 32.7…    073
4  POLYGON ((-117.37213 33.20012, -117.36902 33.2…    073

[5 rows x 195 columns]
```

0.0.2 2. Create a second GeoDataFrame for the file: ~/data/shared/tims/sdcounty_2020_fatal.geojson This data records fatal traffic collisions in San Diego County from https://tims.berkeley.edu.

[18]:
```
sd_fatal_collisions = gpd.read_file('~/data/shared/tims/sdcounty_2020_fatal.
↪geojson')
```

[19]:
```
sd_fatal_collisions.head()
```

[19]:
```
  COLLISION_DATE  COLLISION_TIME  DAY_OF_WEEK  NUMBER_KILLED ALCOHOL_INVOLVED  \
0     2020-12-04            1839            5              1                N
1     2020-12-17            1939            4              1                N
2     2020-12-20            2151            7              2                Y
3     2020-03-06             804            5              1                N
4     2020-05-05            2028            2              4                Y
```

```
                      geometry
0  POINT (-117.08523 33.08923)
1  POINT (-117.07560 33.10338)
2  POINT (-117.09288 33.16100)
3  POINT (-117.06862 33.12628)
4  POINT (-117.06314 33.12370)
```

0.1 3. What are the beginning and ending dates for the collision data?

December 4, 2020 - May 5, 2020

0.2 4. How many fatal collisions occured during this period?

5 fatal collisions occured during this time.

0.3 5. How many individuals were killed in these collisions?

9 people were killed from the collisions.

0.4 6. What percentage of collisions were alcohol involved?

40% of the collisions involved alcohol.

0.5 7. What percentage of the fatalities were alcohol involved?

66.7% of the fatalities involved alcohol.

[22]:
```
import matplotlib.pyplot as plt
```

[24]:
```
san_diego_tracts.plot(ax=ax, color='whitesmoke', edgecolor='black', linewidth=0.
↪5)
```

[24]:
```
<Axes: >
```

```
<Figure size 640x480 with 0 Axes>
```

[25]:
```
sd_fatal_collisions.plot(ax=ax, marker='o', color='red', markersize=5)
```

[25]:
```
<Axes: >
```

```
<Figure size 640x480 with 0 Axes>
```

[26]:
```
ax.set_title('San Diego County - Census Tracts and Fatal Collisions')
plt.show()
```

[27]:
```
plt.show()
```

[28]:
```
print(san_diego_tracts.crs)
```

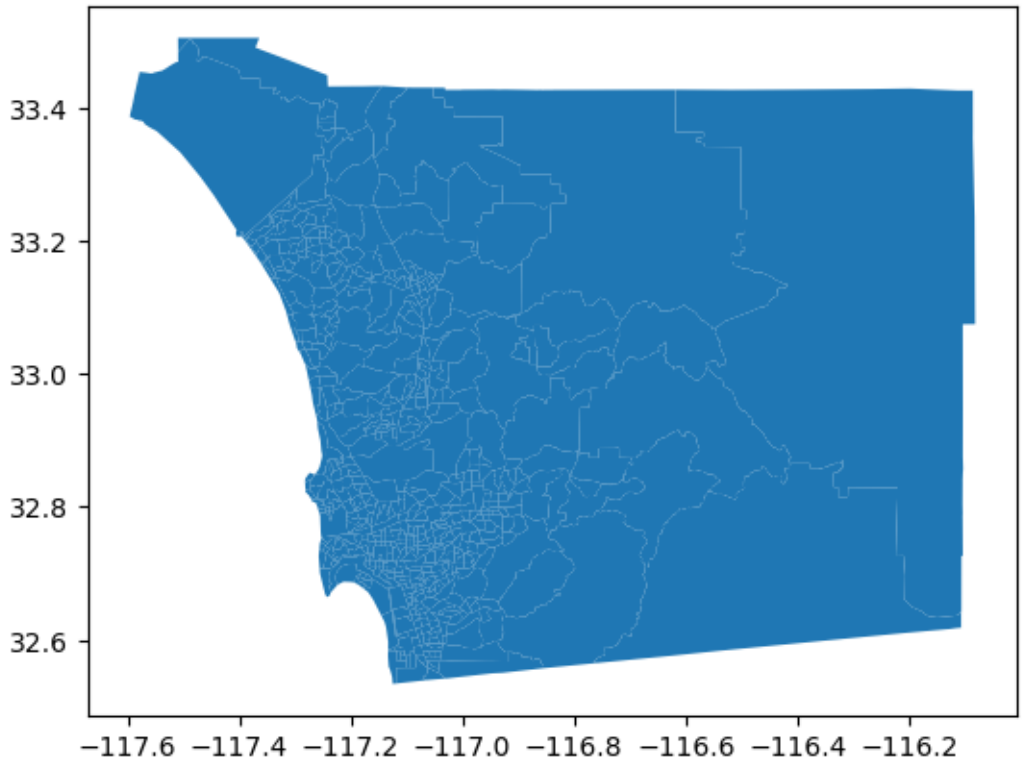
{"$schema": "https://proj.org/schemas/v0.7/projjson.schema.json", "type": "GeographicCRS", "name": "WGS 84", "datum_ensemble": {"name": "World Geodetic System 1984 ensemble", "members": [{"name": "World Geodetic System 1984 (Transit)"}, {"name": "World Geodetic System 1984 (G730)"}, {"name": "World Geodetic System 1984 (G873)"}, {"name": "World Geodetic System 1984 (G1150)"}, {"name": "World Geodetic System 1984 (G1674)"}, {"name": "World Geodetic System 1984 (G1762)"}, {"name": "World Geodetic System 1984 (G2139)"}], "ellipsoid": {"name": "WGS 84", "semi_major_axis": 6378137, "inverse_flattening": 298.257223563}, "accuracy": "2.0", "id": {"authority": "EPSG", "code": 6326}}, "coordinate_system": {"subtype": "ellipsoidal", "axis": [{"name": "Geodetic latitude", "abbreviation": "Lat", "direction": "north", "unit": "degree"}, {"name": "Geodetic longitude", "abbreviation": "Lon", "direction": "east", "unit": "degree"}]}, "scope": "Horizontal component of 3D system.", "area": "World.", "bbox": {"south_latitude": -90, "west_longitude": -180, "north_latitude": 90, "east_longitude": 180}, "id": {"authority": "EPSG", "code": 4326}}

[29]:
```
print(sd_fatal_collisions.crs)
```

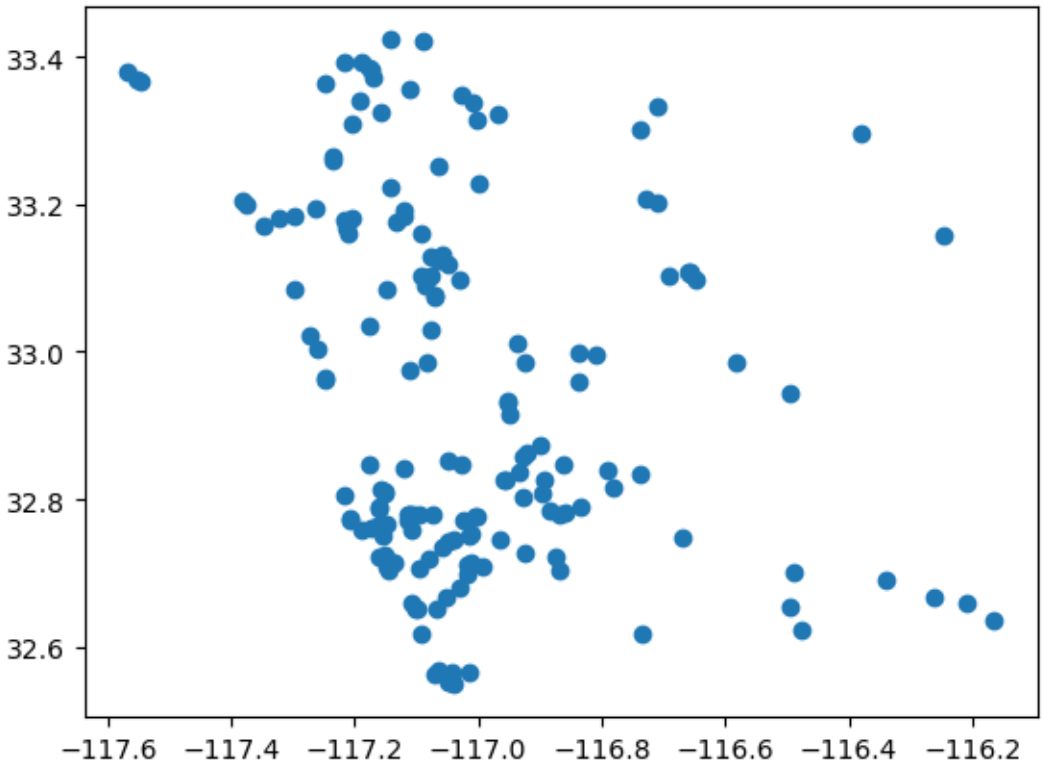
```
EPSG:4326
```

[30]:
```
if san_diego_tracts.crs != sd_fatal_collisions.crs:
    sd_fatal_collisions = sd_fatal_collisions.to_crs(san_diego_tracts.crs)
```

[31]:
```
san_diego_tracts.plot()
plt.show()
```



[32]:
```
sd_fatal_collisions.plot()
plt.show()
```



[33]:
```
import matplotlib.pyplot as plt
```

[36]:
```
sd_fatal_collisions.plot(ax=ax, marker='o', color='red', markersize=5)
```

[36]: <Axes: >

<Figure size 640x480 with 0 Axes>

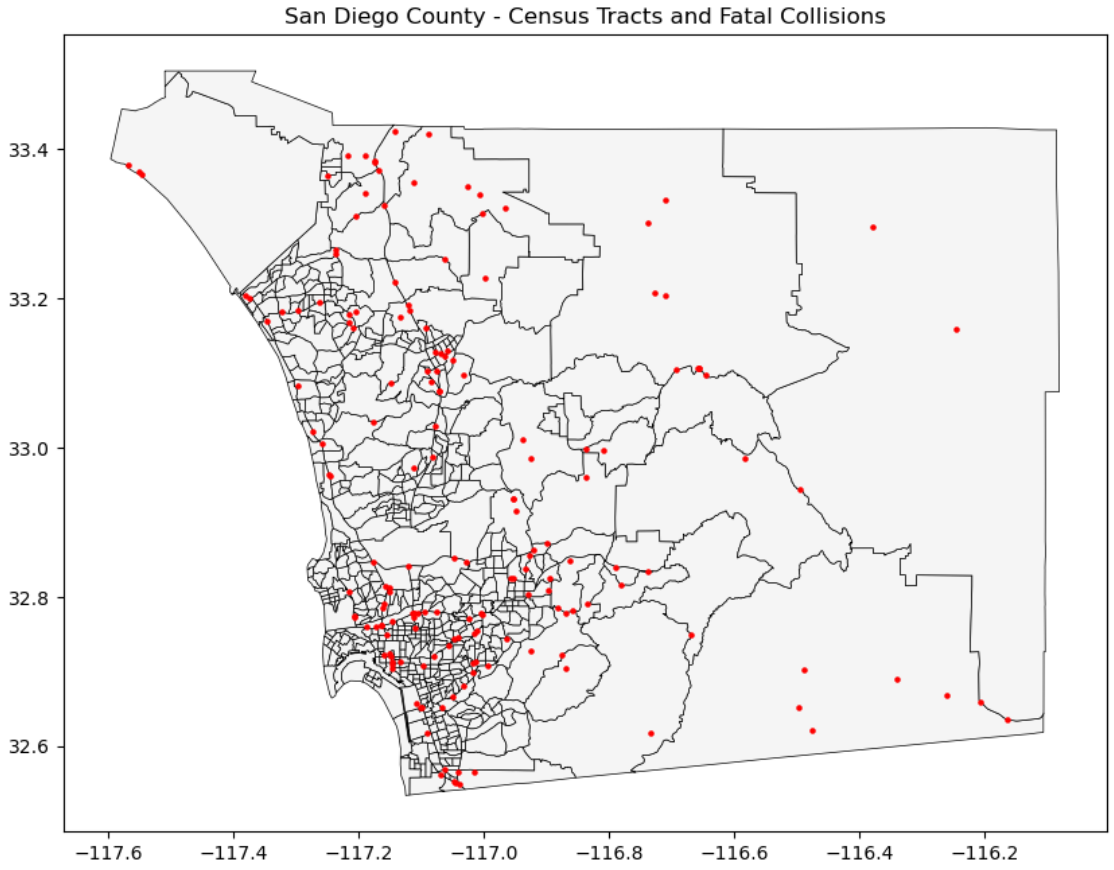
0.5.1 8. Create a two layer map that overlays the collision data on the tracts for San Diego County.

[37]:
```
ax.set_title('San Diego County - Census Tracts and Fatal Collisions')
plt.show()
```

[38]:
```
plt.show()
```

[39]:
```
fig, ax = plt.subplots(figsize=(10, 10))
san_diego_tracts.plot(ax=ax, color='whitesmoke', edgecolor='black', linewidth=0.
↪5)
sd_fatal_collisions.plot(ax=ax, marker='o', color='red', markersize=5)
ax.set_title('San Diego County - Census Tracts and Fatal Collisions')
```

[39]: Text(0.5, 1.0, 'San Diego County - Census Tracts and Fatal Collisions')



[4]:
```
import geopandas as gpd
```

[5]:
```
import matplotlib.pyplot as plt
```

[6]:
```
import pandas as pd
```

[7]:
```
import mapclassify
```

[8]:
```
scag_region = gpd.read_parquet("~/data/scag_region.parquet")
```

[10]:
```
scag_region = gpd.read_parquet('data/scag_region.parquet')
```

[13]:
```
gdf = gpd.read_parquet("~/data/scag_region.parquet")
```

[9]:
```
scag_region['06073'] = scag_region['geoid'].str.slice(2, 5)
```

[10]:
```
san_diego_tracts = scag_region[scag_region['06073'] == '073'].
↪reset_index(drop=True)
```

[13]:
```
sd_fatal_collisions = gpd.read_file('~/data/shared/tims/sdcounty_2020_fatal.
↪geojson')
```

[3]:
```
import geopandas as gpd
```

[8]:
```
scag_region = gpd.read_parquet("~/data/scag_region.parquet")
```

[9]:
```
scag_region['06073'] = scag_region['geoid'].str.slice(2, 5)
```

[10]:
```
san_diego_tracts = scag_region[scag_region['06073'] == '073'].
↪reset_index(drop=True)
```

[13]:
```
sd_fatal_collisions = gpd.read_file('~/data/shared/tims/sdcounty_2020_fatal.
↪geojson')
```

[14]:
```
joined = gpd.sjoin(san_diego_tracts, sd_fatal_collisions, how="left",␣
↪predicate="contains")
```

[16]:
```
joined = gpd.sjoin(san_diego_tracts, sd_fatal_collisions, how="left",␣
↪predicate="contains")
```

[17]:
```
print(joined.columns)
```

```
Index(['geoid', 'n_asian_under_15', 'n_black_under_15', 'n_hispanic_under_15',
       'n_native_under_15', 'n_white_under_15', 'n_persons_under_18',
       'n_asian_over_60', 'n_black_over_60', 'n_hispanic_over_60',
       …
       'p_asian_over_60', 'p_disabled', 'geometry', '06073', 'index_right',
       'COLLISION_DATE', 'COLLISION_TIME', 'DAY_OF_WEEK', 'NUMBER_KILLED',
       'ALCOHOL_INVOLVED'],
      dtype='object', length=201)
```

[18]:
```
collision_counts = joined.groupby('geoid').size()
```

0.5.2 10. Create a new variable in the tracts GeoDataFrame called fatal that reports the number of fatal collisions that occurred in the tract.

[19]:
```
san_diego_tracts['fatal'] = san_diego_tracts['geoid'].map(collision_counts).
↪fillna(0)
```

0.5.3 11. Create a new variable in the collisions GeoDataFrame called tractidx that reports the index of the tract that contains the collision.

[48]:
```
sd_fatal_collisions['tractidx'] = joined['index_right']
```

[24]:
```
joined.reset_index(inplace=True)
```

[25]:
```
sd_fatal_collisions['tract_geoid'] = joined['geoid']
```

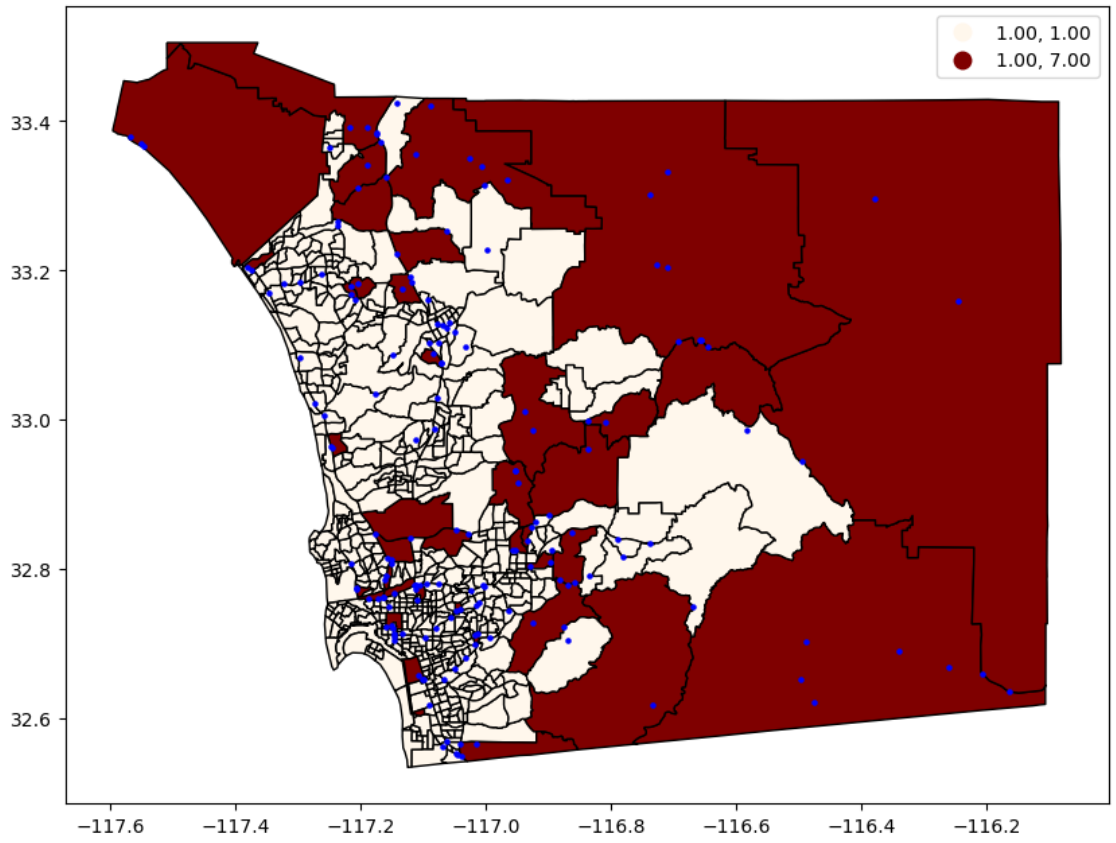
[27]:
```
san_diego_tracts.reset_index(inplace=True, drop=False)
```

[28]:
```
joined = gpd.sjoin(san_diego_tracts, sd_fatal_collisions, how="left",␣
↪predicate="contains")
```

[30]:
```
import matplotlib.pyplot as plt
```

[31]:
```
fig, ax = plt.subplots(figsize=(10, 10))
san_diego_tracts.plot(ax=ax, column='fatal', legend=True, cmap='OrRd',␣
↪scheme='quantiles', edgecolor='black')
sd_fatal_collisions.plot(ax=ax, marker='o', color='blue', markersize=5)
plt.show()
```

```
/opt/tljh/user/lib/python3.10/site-packages/mapclassify/classifiers.py:1592:
UserWarning: Not enough unique values in array to form 5 classes. Setting k to
2.
  self.bins = quantile(y, k=k)
```



0.5.4 12. Create a static, two layer map with the first layer offering a choropleth of variable fatal for each tract, and the second layer showing the locations of the fatal collisions.

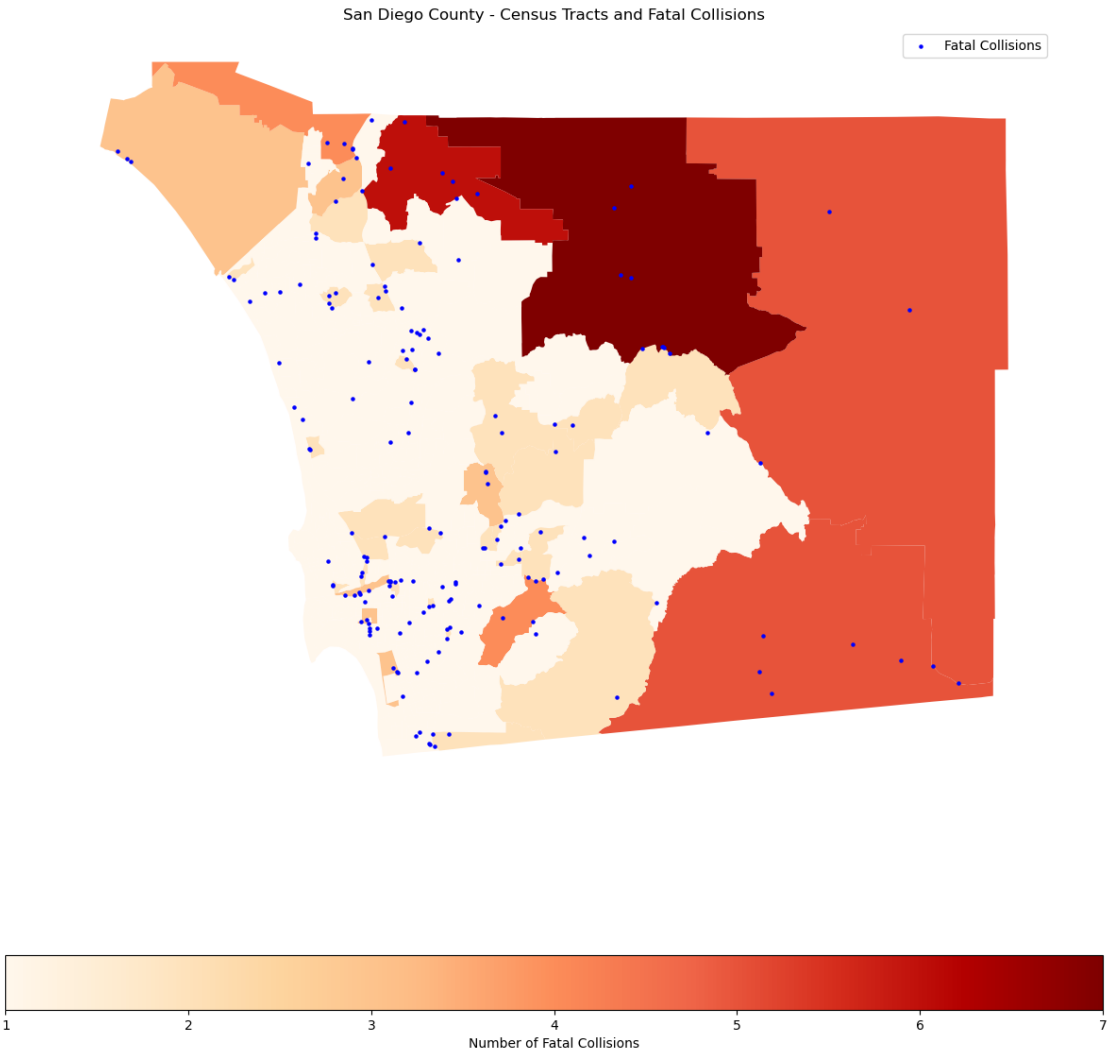
[37]:
```
fig, ax = plt.subplots(figsize=(15, 15))

san_diego_tracts.plot(ax=ax, column='fatal', cmap='OrRd', legend=True,
                      legend_kwds={'label': "Number of Fatal Collisions",
                                   'orientation': "horizontal"})

sd_fatal_collisions.plot(ax=ax, marker='o', color='blue', markersize=5,␣
 ↪label='Fatal Collisions')

ax.set_title('San Diego County - Census Tracts and Fatal Collisions')
ax.set_axis_off()

plt.legend()
plt.show()
```



0.5.5 13. Create an interactive map with the same two layers. Modify the tooltip argument in both layers to limit the number of variables that show when hovering over the polygons/points. The tooltip for the tracts should include the geoid and fatal variables, the tooltip for the point layer should include tractidx as well as the other attributes of the collision.

[39]:
```
import folium
```

[40]:
```
san_diego_center = [32.7157, -117.1611]
```

[41]:
```
m = folium.Map(location=san_diego_center, zoom_start=10)
```

[42]:
```
folium.Choropleth(
    geo_data=san_diego_tracts.to_json(),
```

```
    name='Choropleth',
    data=san_diego_tracts,
    columns=['geoid', 'fatal'],
    key_on='feature.properties.geoid',
    fill_color='YlGn',
    fill_opacity=0.7,
    line_opacity=0.2,
    legend_name='Fatal Collisions Count',
    highlight=True
).add_to(m)
```

[42]:
```
<folium.features.Choropleth at 0x76fe9f1efca0>
```

[47]:
```
joined = gpd.sjoin(sd_fatal_collisions, san_diego_tracts, how="left",␣
↪predicate="intersects")
```

[49]:
```
for _, row in sd_fatal_collisions.iterrows():
    folium.Marker(
        location=[row.geometry.y, row.geometry.x],
        popup=f'Tract ID: {row["tractidx"]}',
        icon=folium.Icon(color='red', icon='info-sign')
    ).add_to(m)
```

[50]:
```
m
```

[50]:
```
<folium.folium.Map at 0x76fe940ca890>
```

0.5.6 15. Based on your exploration, provide an interpretation of the spatial distribution of fatal collisions.

The fatal accidents are noticeably in more densely populated areas, even though the census tracts such as the Santa Ynez Indian Reservation show more fatal collisions, which is due to the size of the tiny census tracts of the urban areas of San Diego.

[ ]:
```

```