PS4 Spatial

Claudia Felipe and Francesca Leon

PS4: Due Sat Nov 2 at 5:00PM Central. Worth 100 points.

1. This problem set is a paired problem set.
2. Play paper, scissors, rock to determine who goes first. Call that person Partner 1. • Partner 1 (name and cnet ID): Francesca Leon (francescaleon) • Partner 2 (name and cnet ID): Claudia Felipe (claudiafelipe)
3. Partner 1 will accept the ps4 and then share the link it creates with their partner. You can only share it with one partner so you will not be able to change it after your partner has accepted.
4. "This submission is our work alone and complies with the 30538 integrity policy." Add your initials to indicate your agreement: **CF FL**
5. "I have uploaded the names of anyone else other than my partner and I worked with on the problem set here" (1 point)
6. Late coins used this pset: **0** Late coins left after submission: **4**

Style Points (10 pts)

Submission Steps (10 pts)

Download and explore the Provider of Services (POS) file (10 pts)

1. The variables I pulled are: PRVDR_CTGRY_SBTYP_CD, PRVDR_CTGRY_CD, FAC_NAME, PRVDR_NUM, STATE_CD, PGM_TRMNTN_CD, ZIP_CD.

```
import pandas as pd
import altair as alt
import time
import numpy as np
import warnings
import geopandas as gpd
import os
```

```
warnings.filterwarnings('ignore')
#alt.renderers.enable('png')
alt.data_transformers.disable_max_rows()

base_directory = 'C:/Users/clfel/Documents/GitHub/problem-set-4-fran-clau'
```

2.

a. The number of hospitals reported are 7,245. This number doesn't make full sense because seems to be too high.

```
pos2016 = pd.read_csv(os.path.join(base_directory, 'pos2016.csv'))
pos2016 = pos2016[(pos2016['PRVDR_CTGRY_CD']==1) &
 ↪  (pos2016['PRVDR_CTGRY_SBTYP_CD']==1)]
pos2016['year'] = 2016

pos2016['PRVDR_NUM'].nunique()
```

```
7245
```

b. Based on the American Hospital Association Statistics, the number of short term hospitals in 2018 (oldest available report) was 4,840. This number differs significantly from the number calculated using the database (7,245). This could happen because the hospitals are mistakenly categorized as short-term in the database. Another explanation could be that some of the short-term hospitals are closed but still appear in the database.

3.

```
pos2017 = pd.read_csv(os.path.join(base_directory, 'pos2017.csv'))
pos2017 = pos2017[(pos2017['PRVDR_CTGRY_CD']==1) &
 ↪  (pos2017['PRVDR_CTGRY_SBTYP_CD']==1)]
pos2017['year'] = 2017

pos2018 = pd.read_csv(os.path.join(base_directory, 'pos2018.csv'),
 ↪  encoding="ISO-8859-1")
pos2018 = pos2018[(pos2018['PRVDR_CTGRY_CD']==1) &
 ↪  (pos2018['PRVDR_CTGRY_SBTYP_CD']==1)]
pos2018['year'] = 2018
```

```
pos2019 = pd.read_csv(os.path.join(base_directory, 'pos2019.csv'),
 ↪  encoding="ISO-8859-1")
pos2019 = pos2019[(pos2019['PRVDR_CTGRY_CD']==1) &
 ↪  (pos2019['PRVDR_CTGRY_SBTYP_CD']==1)]
pos2019['year'] = 2019

pos = pd.concat([pos2016, pos2017, pos2018, pos2019], ignore_index=True)


alt.Chart(pos, title = 'Number of Observations per Year').mark_bar().encode(
  alt.X('year:O').title('Year'),
  alt.Y('count():Q').title('Number of Observations')
).properties(width = 500)
```

```
alt.Chart(...)
```

4.

a.

```
alt.Chart(pos, title = 'Number of Unique Hospitals per
 ↪  Year').mark_bar().encode(
  alt.X('year:O').title('Year'),
  alt.Y('distinct(PRVDR_NUM):Q').title('Number of Unique Hospitals')
).properties(width = 500)
```

```
alt.Chart(...)
```

b. The graphs are the same, with the same numbers. This means that the database contains the same number of observations and unique hospitals. This tells that the structure of the data is attached to the unique hospitals, those being the unit of observation.

Identify hospital closures in POS file (15 pts) (*)

1. The total suspected hospital closures is 174.

```
# Filter active hospitals in 2016
active_2016 = pos2016[pos2016['PGM_TRMNTN_CD'] == 0][['PRVDR_NUM',
 ↪  'FAC_NAME', 'ZIP_CD']]

# Create a list to store the suspected closure year
closure_data = []

# Iterate over each hospital active in 2016 and check if it remains active in
 ↪  the following years
for _, row in active_2016.iterrows():
    provider_id = row['PRVDR_NUM']
    facility_name = row['FAC_NAME']
    zip_code = row['ZIP_CD']

    # Check each successive year
    if provider_id not in pos2017[pos2017['PGM_TRMNTN_CD'] ==
     ↪  0]['PRVDR_NUM'].values:
        closure_data.append([facility_name, zip_code, 2017])
    elif provider_id not in pos2018[pos2018['PGM_TRMNTN_CD'] ==
     ↪  0]['PRVDR_NUM'].values:
        closure_data.append([facility_name, zip_code, 2018])
    elif provider_id not in pos2019[pos2019['PGM_TRMNTN_CD'] ==
     ↪  0]['PRVDR_NUM'].values:
        closure_data.append([facility_name, zip_code, 2019])

# Convert the results into a DataFrame
closure_df = pd.DataFrame(closure_data, columns=['FAC_NAME', 'ZIP_CD', 'Year
 ↪  of Susp. Closure'])

print("Total suspected hospital closures:", closure_df.shape[0])
```

```
Total suspected hospital closures: 174
```

2.

```
# Sort hospitals by name and select the first 10 rows
sorted_hospitals = closure_df.sort_values(by='FAC_NAME')[['FAC_NAME', 'Year
 ↪  of Susp. Closure']].head(10)

# Display the results
sorted_hospitals.style.hide()
```

Table 1

FAC_NAME	Year of Susp. Closure
ABRAZO MARYVALE CAMPUS	2017
ADVENTIST MEDICAL CENTER - CENTRAL VALLEY	2017
AFFINITY MEDICAL CENTER	2018
ALBANY MEDICAL CENTER / SOUTH CLINICAL CAMPUS	2017
ALLEGIANCE SPECIALTY HOSPITAL OF KILGORE	2017
ALLIANCE LAIRD HOSPITAL	2019
ALLIANCEHEALTH DEACONESS	2019
ANNE BATES LEACH EYE HOSPITAL	2019
ARKANSAS VALLEY REGIONAL MEDICAL CENTER	2017
BANNER CHURCHILL COMMUNITY HOSPITAL	2017

3.

```
# Count the total number of active hospitals by ZIP code for each year
active_by_zip_2016 = pos2016[pos2016['PGM_TRMNTN_CD'] ==
 ↪  0].groupby('ZIP_CD').size().to_dict()
active_by_zip_2017 = pos2017[pos2017['PGM_TRMNTN_CD'] ==
 ↪  0].groupby('ZIP_CD').size().to_dict()
active_by_zip_2018 = pos2018[pos2018['PGM_TRMNTN_CD'] ==
 ↪  0].groupby('ZIP_CD').size().to_dict()
active_by_zip_2019 = pos2019[pos2019['PGM_TRMNTN_CD'] ==
 ↪  0].groupby('ZIP_CD').size().to_dict()

# Create a DataFrame from the dictionaries
active_counts_df = pd.DataFrame({
    '2016': pd.Series(active_by_zip_2016),
    '2017': pd.Series(active_by_zip_2017),
    '2018': pd.Series(active_by_zip_2018),
    '2019': pd.Series(active_by_zip_2019)
})

# Fill any missing values with 0, in case a ZIP code has no active hospitals
 ↪  in a given year
active_counts_df = active_counts_df.fillna(0).astype(int)

# Create new columns for the difference in active hospital counts between
 ↪  consecutive years
active_counts_df['2017-2016'] = active_counts_df['2017'] -
 ↪  active_counts_df['2016']
```

```
active_counts_df['2018-2017'] = active_counts_df['2018'] -
 ↪  active_counts_df['2017']
active_counts_df['2019-2018'] = active_counts_df['2019'] -
 ↪  active_counts_df['2018']

active_counts_df = active_counts_df.reset_index()
active_counts_df = active_counts_df.rename(columns={'index': 'ZIP_CD'})

# Perform the merge on 'ZIP_CD'
merged_df = pd.merge(closure_df, active_counts_df, on='ZIP_CD', how='left')

#
merged_df['change'] = np.where(
    merged_df['Year of Susp. Closure'] == 2017, merged_df['2017-2016'],
    np.where(
        merged_df['Year of Susp. Closure'] == 2018, merged_df['2018-2017'],
        np.where(
            merged_df['Year of Susp. Closure'] == 2019,
 ↪  merged_df['2019-2018'],
            np.nan
        )
    )
)
```

a. The number of hospitals potentially merged/acquired is 8

```
# Possible mergers or acquisitions: hospitals with "change" <= 0
merged_or_acquisition = merged_df[(merged_df['change'] >= 0) &
 ↪  (merged_df['change'].notna())]
n_merged = merged_or_acquisition.shape[0]
print("Number of hospitals potentially merged/acquired:", n_merged)
```

```
Number of hospitals potentially merged/acquired: 8
```

b.

```
# Remaining valid closures with "change" >= 1
df_corrected = merged_df[(merged_df['change'] < 0) &
 ↪  (merged_df['change'].notna())]
print("Remaining valid closures after excluding mergers/acquisitions:",
 ↪  df_corrected.shape[0])
```

```
Remaining valid closures after excluding mergers/acquisitions: 166
```

c. The remaining valid closures after excluding mergers/acquisitions is 166.

```
# Sort the corrected DataFrame by hospital name and display the first 10 rows

sorted_corrected_closures =
 ↪  df_corrected.sort_values(by='FAC_NAME')[['FAC_NAME', 'ZIP_CD', 'Year of
 ↪  Susp. Closure']].head(10)
sorted_corrected_closures['ZIP_CD']=sorted_corrected_closures['ZIP_CD'].astype(int)
sorted_corrected_closures.style.hide()
```

Table 2

FAC_NAME	ZIP_CD	Year of Susp. Closure
ABRAZO MARYVALE CAMPUS	85031	2017
ADVENTIST MEDICAL CENTER - CENTRAL VALLEY	93230	2017
AFFINITY MEDICAL CENTER	44646	2018
ALBANY MEDICAL CENTER / SOUTH CLINICAL CAMPUS	12208	2017
ALLEGIANCE SPECIALTY HOSPITAL OF KILGORE	75662	2017
ALLIANCE LAIRD HOSPITAL	39365	2019
ALLIANCEHEALTH DEACONESS	73112	2019
ANNE BATES LEACH EYE HOSPITAL	33136	2019
ARKANSAS VALLEY REGIONAL MEDICAL CENTER	81050	2017
BANNER CHURCHILL COMMUNITY HOSPITAL	89406	2017

Download Census zip code shapefile (10 pt)

1. a.
   b.
2.

Calculate zip code's distance to the nearest hospital (20 pts) (*)

1.
2.
3.
4. a.
   b.
   c.

5. a.
   b.
   c.

Effects of closures on access in Texas (15 pts)

1.
2.
3.
4.

Reflecting on the exercise (10 pts)