

PS4 v1.1: Spatial

Boya Lin and Zidan Kong

PS4: Due Sat Nov 2 at 5:00PM Central. Worth 100 points. We use (*) to indicate a problem that we think might be time consuming.

Style Points (10 pts)

Please refer to the minilesson on code style [here](#).

Submission Steps (10 pts)

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): Boya Lin, boyal
 - Partner 2 (name and cnet ID):
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: ****BL** **__****
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: ****__**** Late coins left after submission: ****__****
7. Knit your **ps4.qmd** to an PDF file to make **ps4.pdf**,
 - The PDF should not be more than 25 pages. Use **head()** and re-size figures when appropriate.
8. (Partner 1): push **ps4.qmd** and **ps4.pdf** to your github repo.
9. (Partner 1): submit **ps4.pdf** via Gradescope. Add your partner on Gradescope.
10. (Partner 1): tag your submission in Gradescope

Important: Repositories are for tracking code. **Do not commit the data or shapefiles to your repo.** The best way to do this is with `.gitignore`, which we have covered in class. If you do accidentally commit the data, Github has a [guide](#). The best course of action depends on whether you have pushed yet. This also means that both partners will have to download the initial raw data and any data cleaning code will need to be re-run on both partners' computers.

```
import pandas as pd
import altair as alt
alt.renderers.enable("png")
```

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

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

1.

Variables we pulled in `pos2016.csv` include `PRVDR_CTGRY_SBTYP_CD`, `PRVDR_CTGRY_CD`, `CITY_NAME`, `FAC_NAME` (Facility Name), `ORGNL_PRTCPTN_DT`, `PRVDR_NUM` (CMS Certification Number), `PGM_TRMNTN_CD` (Termination Code), `TRMNTN_EXPRTN_DT`, and `ZIP_CD`.

2. a.

```
pos_2016 = pd.read_csv('pos2016.csv')
short_term_hospitals_2016 = pos_2016[(pos_2016['PRVDR_CTGRY_CD'] == 1) & (
    pos_2016['PRVDR_CTGRY_SBTYP_CD'] == 1)].copy()
# ChatGPT reference for SettingWithCopyWarning, the usage of .copy

count_2016 = short_term_hospitals_2016.shape[0]
print(f'Number of short-term hospitals in 2016: {count_2016}')
```

Number of short-term hospitals in 2016: 7245

This number appears slightly high, since the American Hospital Association (AHA) reported a total of 6,120 hospitals in the U.S. in 2024. In consideration of the growing hospital industry, we would generally expect the number of hospitals in 2016 to be lower than that in 2024.

Data Source (AHA): <https://www.aha.org/statistics/fast-facts-us-hospitals>

b.

Unlike 7,245 hospitals reported in our data, the American Hospital Association (AHA) reports a total number of 5,534 U.S. hospitals for 2016. The difference may arise from factors such as hospitals name changes, acquisitions, or the presence of branches or divisions, which can lead to multiple entries for the same facility in our data. As a result, when using `.nunique()`, these hospitals/facilities might be counted separately, leading to an inflated total.

Data Source (AHA): <chrome-extension://efaidnbmnnnibpcjpcglclefindmkaj/https://www.aha.org/system/files/2018-01/Fast%20Facts%202018%20pie%20charts.pdf>

3.

```
pos_2017 = pd.read_csv('pos2017.csv')
short_term_hospitals_2017 = pos_2017[(pos_2017['PRVDR_CTGRY_CD'] == 1) & (
    pos_2017['PRVDR_CTGRY_SBTYP_CD'] == 1)].copy()

pos_2018 = pd.read_csv('pos2018.csv', encoding='ISO-8859-1')
short_term_hospitals_2018 = pos_2018[(pos_2018['PRVDR_CTGRY_CD'] == 1) & (
    pos_2018['PRVDR_CTGRY_SBTYP_CD'] == 1)].copy()

pos_2019 = pd.read_csv('pos2019.csv', encoding='ISO-8859-1')
short_term_hospitals_2019 = pos_2019[(pos_2019['PRVDR_CTGRY_CD'] == 1) & (
    pos_2019['PRVDR_CTGRY_SBTYP_CD'] == 1)].copy()
# ChatGPT reference for "UnicodeDecodeError: 'utf-8' codec can't decode byte
↳ 0xa0 in position 119424", solution: add encoding='ISO-8859-1'

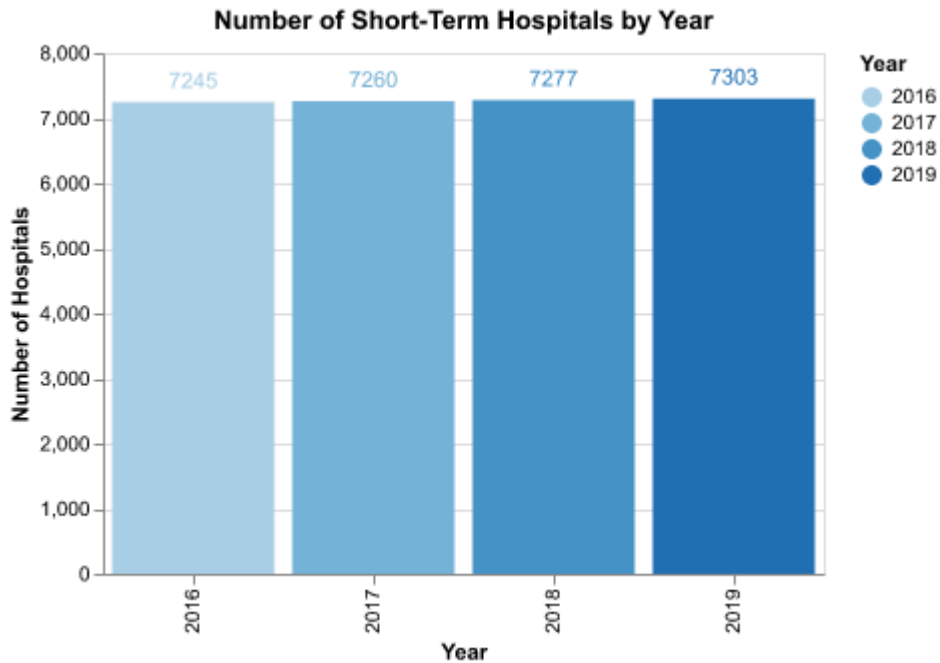
short_term_hospitals_2016.loc[:, 'Year'] = 2016
short_term_hospitals_2017.loc[:, 'Year'] = 2017
short_term_hospitals_2018.loc[:, 'Year'] = 2018
short_term_hospitals_2019.loc[:, 'Year'] = 2019
all_hospitals = pd.concat([short_term_hospitals_2016,
                           short_term_hospitals_2017,
                           short_term_hospitals_2018,
                           short_term_hospitals_2019], ignore_index=True)
# Pandas official website reference for pd.concat:
↳ https://pandas.pydata.org/docs/reference/api/pandas.concat.html

hospitals_by_year = all_hospitals.groupby(
    'Year').size().reset_index(name='Number of Hospitals')

bar_chart = alt.Chart(hospitals_by_year).mark_bar().encode(
    x='Year:O', y='Number of Hospitals:Q', color='Year:O'
).properties(
    title='Number of Short-Term Hospitals by Year', width=360, height=260
```

```
)

text_labels = bar_chart.mark_text(
    align='center', baseline='bottom', dy=-5).encode(text='Number of
↪ Hospitals:Q')
hospital_chart = bar_chart + text_labels
hospital_chart
# ChatGDP reference for adding exact values of data above bar chart
```



4. a.

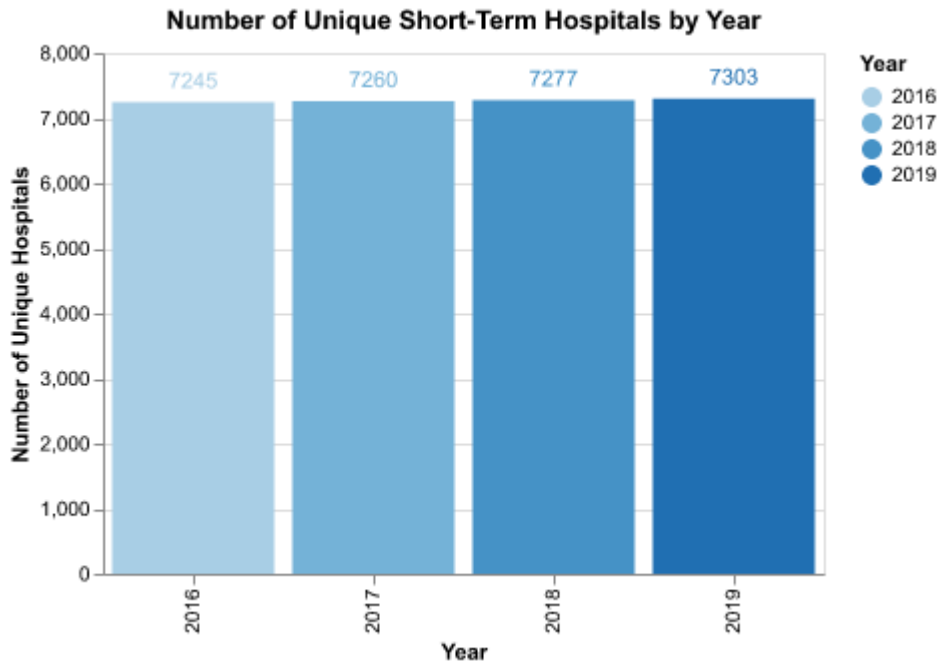
```
unique_hospitals_by_year = all_hospitals.groupby(
    'Year')['PRVDR_NUM'].nunique().reset_index(name='Number of Unique
↪ Hospitals')

bar_chart = alt.Chart(unique_hospitals_by_year).mark_bar().encode(
    x='Year:O', y='Number of Unique Hospitals:Q', color='Year:O'
).properties(
    title='Number of Unique Short-Term Hospitals by Year',
    width=360, height=260
)
```

```

text_labels = bar_chart.mark_text(
    align='center', baseline='bottom', dy=-5).encode(text='Number of Unique
↪ Hospitals:Q')
unique_hospitals_chart = bar_chart + text_labels
unique_hospitals_chart

```



b.

Both plots present the same counts of observations for each year. It suggests that each observation in our dataset corresponds to a unique hospital (identified by its CMS certification number). In other words, there are no duplicate records for any hospital in any year.

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

- 1.
- 2.
3.
 - a.
 - b.
 - c.

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.
5. a.
 b.
 c.

Effects of closures on access in Texas (15 pts)

- 1.
- 2.
- 3.
- 4.

Reflecting on the exercise (10 pts)