	nearest_distance_miles	closure_count_x	ZIP_CD	closure_count_y
0	0.0	1.0	79520	1
1	0.0	1.0	79529	1
2	0.0	1.0	79902	1
3	0.0	1.0	76645	1
4	0.0	1.0	77065	1
5	0.0	1.0	75140	1
6	0.0	1.0	78336	1
7	0.0	1.0	78613	1
8	0.0	1.0	75235	1
9	0.0	1.0	75051	1
10	0.0	1.0	75087	1
11	0.0	1.0	76520	1
12	0.0	1.0	76531	1
13	0.0	1.0	75390	1

```
The number of directly affected zip codes in Texas 342
```

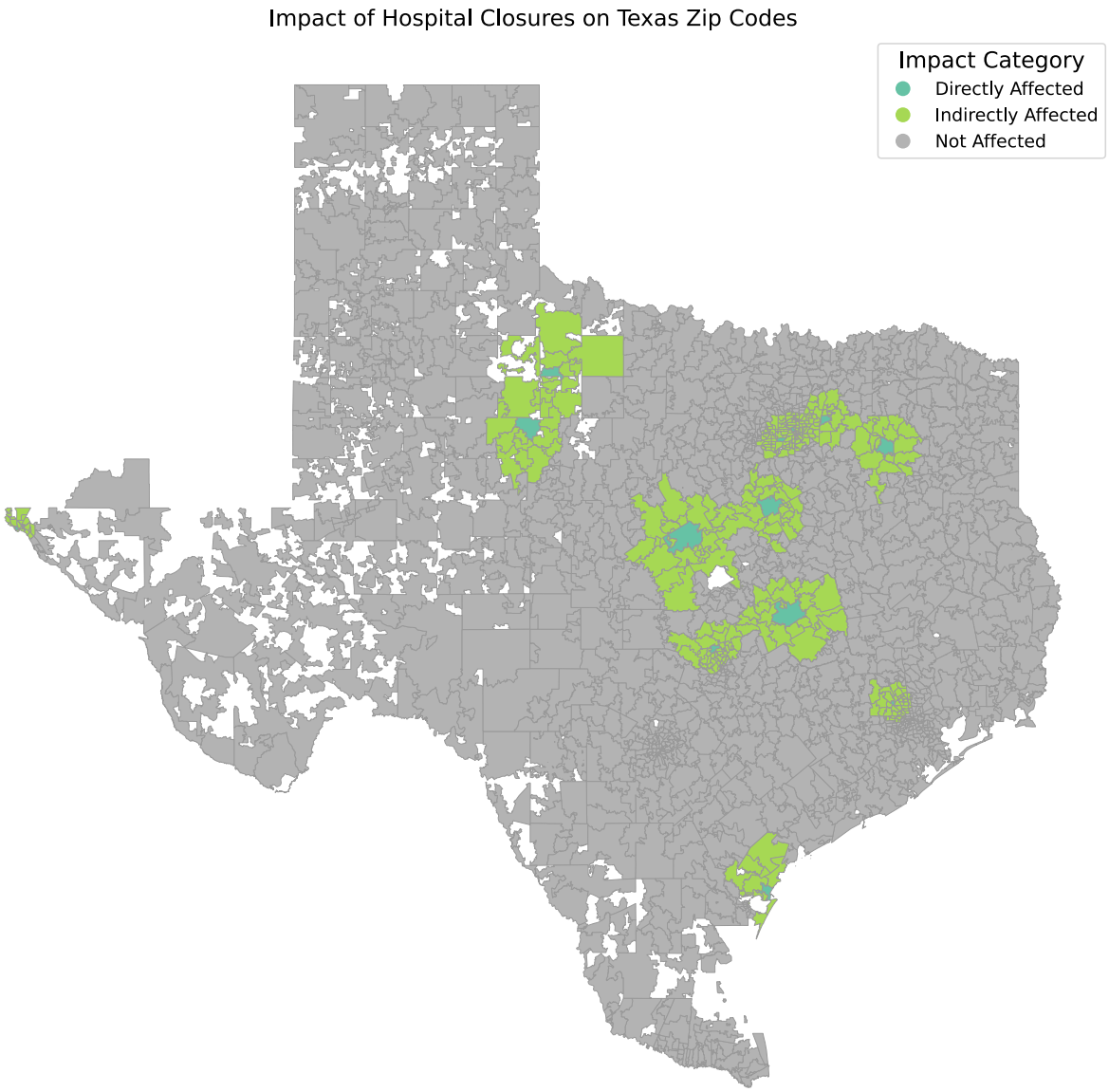
4.

```
zips_texas['impact_category'] = 'Not Affected'

zips_texas.loc[
    zips_texas['ZCTA5'].isin(directly_affected['ZCTA5']), 'impact_category'
] = 'Directly Affected'

zips_texas.loc[
    zips_texas['ZCTA5'].isin(indirectly_affected['ZCTA5_left']),
    'impact_category'
] = "Indirectly Affected"

fig, ax = plt.subplots(figsize=(14, 14))
zips_texas.plot(
    column='impact_category',
    cmap='Set2',
    linewidth=0.1,
    ax=ax,
    edgecolor='0.6',
    legend=True,
    legend_kwds={'title': "Impact Category", 'fontsize': 'large',
    'title_fontsize': '15'}
)
ax.set_title('Impact of Hospital Closures on Texas Zip Codes', fontsize=15)
ax.set_axis_off()
plt.show()
```



Reflecting on the exercise (10 pts)

Partner_1:

The "first-pass" method, which removes suspected closures in zip codes where the number of active hospitals does not decrease the year after the suspected closure, is a helpful starting point. However, it has some limitations.

As mentioned in section2.3, the first-pass method ignores the case where both the closure year and the following year have zero active hospitals, and it incorrectly categorize this situation as a potential merger. To solve this, more considerations are needed like introducing a rule for cases where both the closure year and the following year have zero active hospitals: check if the area had limited healthcare resources. If so, classify it as a true closure rather than a potential merger.

Hospital data might not be fully updated in the following year. Some hospitals may actually be closed while having a late report, leading to an unchanged active hospitals number. There are also cases where hospitals relocate like moving in to or out from a particular zip code area while keeping the same total number. Consolidating and seperating could also happen at the same time while the total number of active hospitals might remain unchanged.

Ways to do better can be tracking hospitals for two to three more years before confirming a closure. This allows enough time for data to be updated and corrected, giving a more accurate picture of real changes. Cross checking CMS and specific address or geographic coordinates rather than just counting hospitals in the zip code or use CMS seperately. By observing exact locations and CMS at same time, we could identify hospitals that have moved or merged.

Partner_2:

ZIP codes with at least one hospital closure between 2016 and 2019 are the most directly impacted outcomes. Here we begin by computationally ignoring follow-up in subsequent years of closure, and also defaulting to the idea that hospital closures within a given ZIP code mean fewer visits for residents.

The current methodology has limitations, such as the fact that distance effectively ignores the population density of the corresponding area in English: zip code areas vary widely in size and population density. Some areas may have a large number of residents, while others are relatively empty. Using zip codes alone to delineate areas may mask the actual distribution of health care needs and resources.

Improvements can be made by considering a combination of factors: for example, when assessing accessibility to hospitals, taking into account the distribution of populations and residential areas, transportation networks, and public transportation options (which are often included in geographic data as well) can provide a more complete picture of how residents actually travel to hospitals.