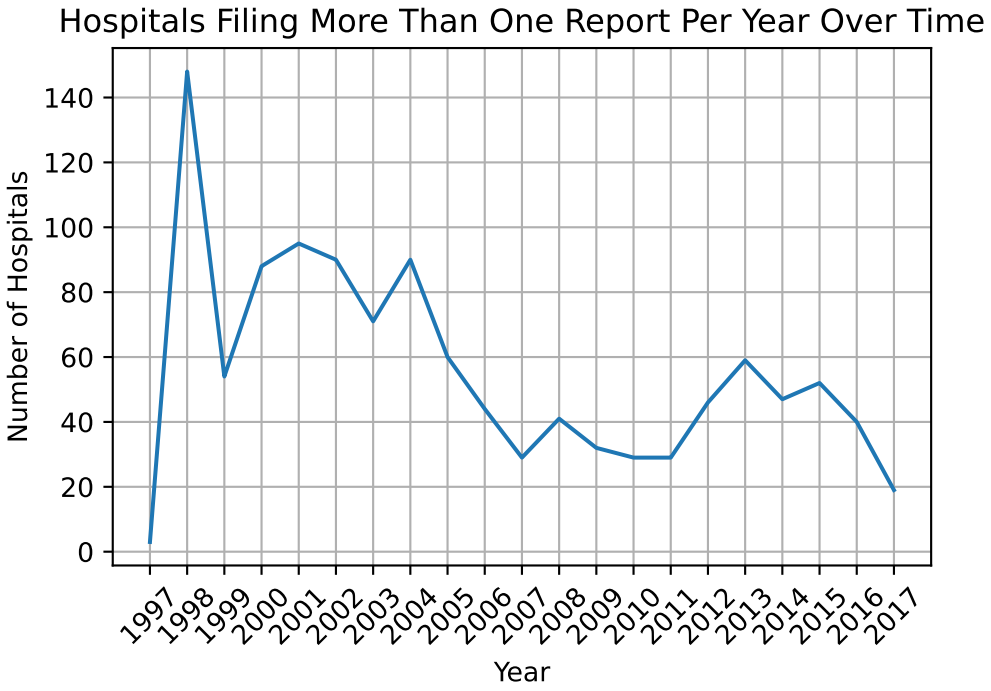
Homework 2

Research Methods, Spring 2025

Ryan Scholte

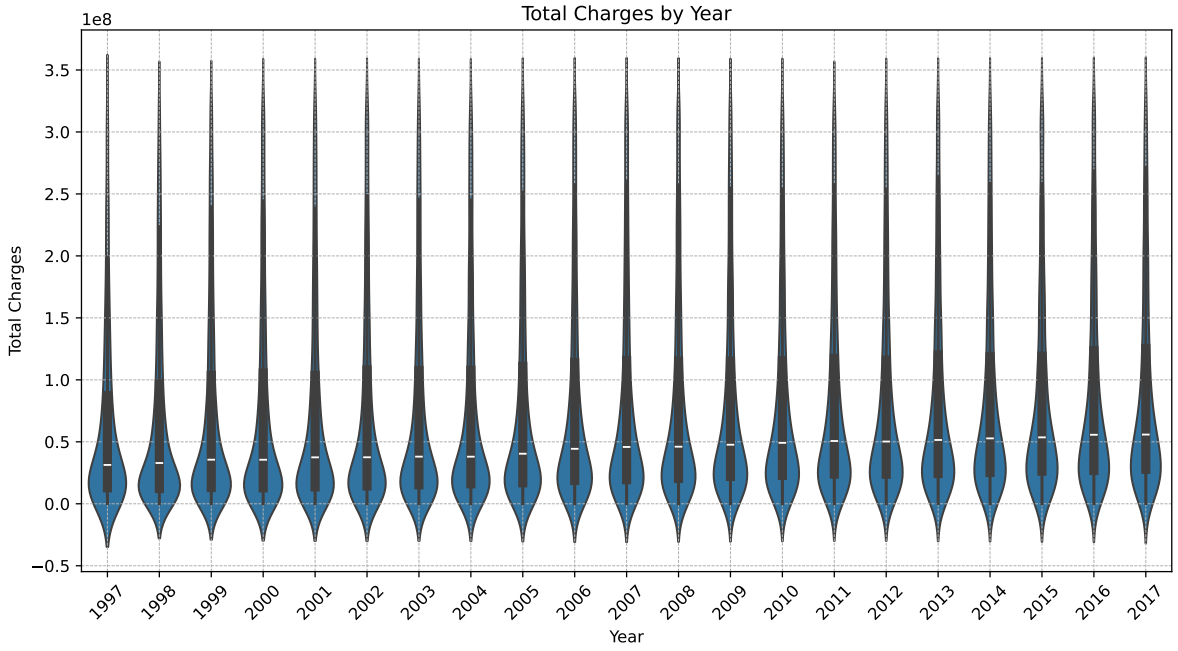
You can access the Repository



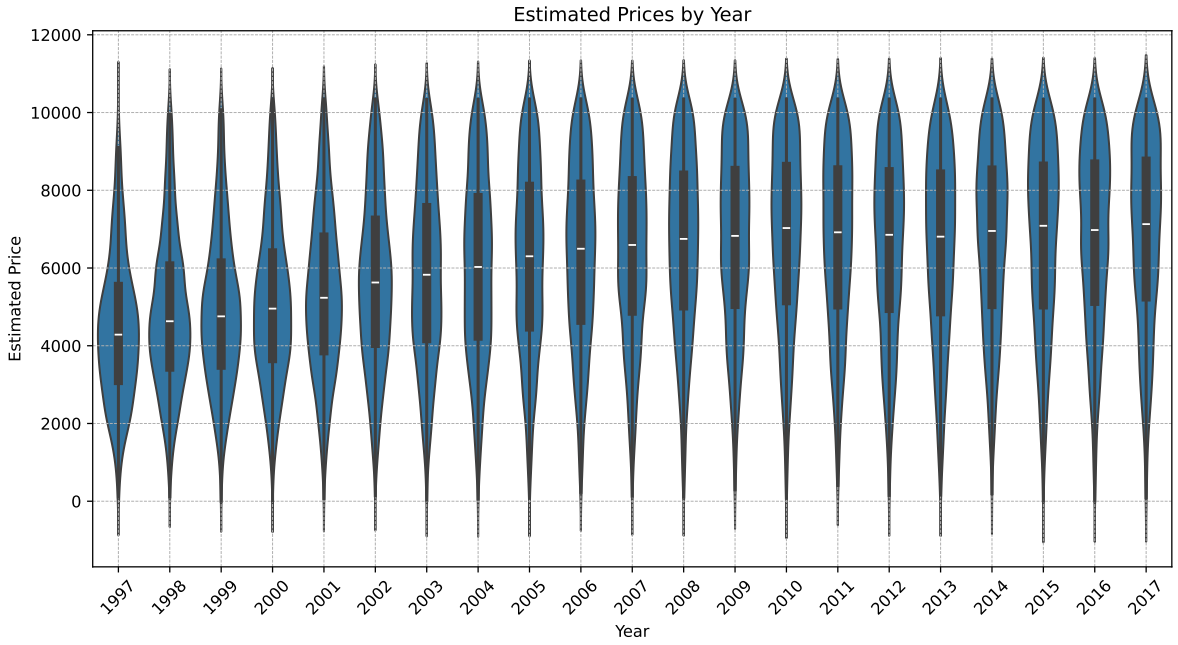
2

```
Number of unique hospital IDs: 9323
```

3


Total Charges by Year
1e8
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
Total Charges
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
Year

4



5

```
Mean price for penalized hospitals: 9496.43
Mean price for non-penalized hospitals: 9699.71
```

6

	Quartile	Treated_Mean_Price	Control_Mean_Price
0	1	6655.93	7853.99
1	2	9418.92	8564.88
2	3	9793.63	9921.73
3	4	11682.48	12503.75

7a

```
/var/folders/mn/l2nrwsxn24g6ywz6ygh2fxp40000gn/T/ipykernel_43207/1105793619.py:11: FutureWarn
  variance_by_quartile = control_df.groupby(covariate)['beds'].var().fillna(1)


ATE using Nearest Neighbor Matching (Inverse Variance Distance): -847.5152

Treatment Effect Estimates: Matching

                    Est.       S.e.          z      P>|z|      [95% Conf. int.]
--------------------------------------------------------------------------------
           ATE   -329.858    343.520     -0.960      0.337  -1003.157    343.441
           ATC   -337.478    343.996     -0.981      0.327  -1011.711    336.755
           ATT   -260.459    341.825     -0.762      0.446   -930.436    409.519

/Users/ryanscholte/anaconda3/lib/python3.10/site-packages/causalinference/estimators/matching
To use the future default and silence this warning we advise to pass `rcond=None`, to keep us
  return np.linalg.lstsq(X, Y)[0][1:]  # don't need intercept coef
```

7b

```
ATE using Nearest Neighbor Matching (Mahalanobis Distance):

Treatment Effect Estimates: Matching

                    Est.       S.e.          z      P>|z|      [95% Conf. int.]
--------------------------------------------------------------------------------
           ATE   -329.858    343.520     -0.960      0.337  -1003.157    343.441
           ATC   -337.478    343.996     -0.981      0.327  -1011.711    336.755
           ATT   -260.459    341.825     -0.762      0.446   -930.436    409.519

/Users/ryanscholte/anaconda3/lib/python3.10/site-packages/causalinference/estimators/matching
To use the future default and silence this warning we advise to pass `rcond=None`, to keep us
  return np.linalg.lstsq(X, Y)[0][1:]  # don't need intercept coef
```

7c

```
ATE using Propensity Score Matching:

Treatment Effect Estimates: Matching

                    Est.       S.e.          z      P>|z|      [95% Conf. int.]
-------------------------------------------------------------------------------
           ATE   -329.858    343.520     -0.960      0.337  -1003.157    343.441
           ATC   -337.478    343.996     -0.981      0.327  -1011.711    336.755
           ATT   -260.459    341.825     -0.762      0.446   -930.436    409.519

/Users/ryanscholte/anaconda3/lib/python3.10/site-packages/causalinference/estimators/matching
To use the future default and silence this warning we advise to pass `rcond=None`, to keep us
  return np.linalg.lstsq(X, Y)[0][1:]  # don't need intercept coef
```

7d

```
ATE using IPW: nan
Linear Regression Coefficients:
{'penalty': -0.005512252615467795, 'beds': 25.854038803013, 'mcaid_discharges': -0.499550200
ATE using Linear Regression: 460.0315
```

8

all ATE are the same

9

no, I dont think my data and code is correct yet. But even with that I think there are many other issues we talked about that we cant assume to determine the causality. those can be unobserved factors like patient acuity that affect the chance of a hospital being penailized. This means penalized hospitals arent assigned randomly and have other factors that are not fully controlled for.

10

It was very difficult especially if the cleaning is not done correctly. Also using a bunch new packages is tough. I learned that file paths are super frustrating and can change all the time but there a bunch of differeent solutions and I haven't found a consistent or one I have learned to use consistently yet.