Final Report
Author: Songru Wang
Data: Dec 10, 2021

Introduction:

Since nowadays there are lot of smokers around the world and many policies have been applied to curb the number of smokers, I would like to figure out whether the number of smokers is actually decreased or not. This data is about "Never smoked trend for 1995-2010". I've found the data from Centers for Disease Control and Prevention website.

Then I want to figure out:
1. Whether non-smokers in each different city become less or not as time goes by
2. The relationship between the change in the number of smokers and the state.

Description of the data:

While analysing the original data table,I found there are 876 obervations, and 7 variables. In this data set, 56 states (including several islands next to the US) are included. And there is no missing values in the data table.

Data Analysis:

According to the four graphs, I found the number of people who smoke everyday shows a decrease as times goes by. The number of people who smoke someday shows a gradual increase as times goes by.The number of people who never smoke becomes more as time goes by.The number of people who is former smokers, in other words, the people who quit smoking, becomes more as times goes by.

Then I divide these 56 states into 5 geographical categories: Northeast, Southwest, West, Southeast, and Midwest. Alaska is hard to define which region it belongs to, so I leave it as NA value.Guam is an organized, unincorporated territory of the United States, so I leave it as NA value.District of Columbia is not a state.
I also choose to not define the geographic region of Virgin Island and Hawaii

Smoke Everyday in each geographic region shows a decreasing tendency. Smoke Somedays in each geographical region shows a slight increase. Former Smoker showed decreasing tendency in the West region, increasing tendency in Northeast, Midwest, and Southeast.

Conclusion:

Based on the analysis, we can conclude that the number of smokers decreases as time goes by, the number of former smokers increases, and the number of people who never smoke increases. This is possibly due to education, policy, or the taxes on cigerattes. However, the changes in different geographical regions is different, which needs further data collection and data analysis.

Table 1:

```
── Data Summary ────────────────────────
                           Values
Name                       smoke
Number of rows             876
Number of columns          7
_______________________
Column type frequency:
  character                2
  numeric                  5
_______________________
Group variables            None

── Variable type: character ────────────────────────────────────────────
  skim_variable n_missing complete_rate   min   max empty n_unique
1 State                 0         1         4    40     0       56
2 Location 1           37         0.958    11    60     0       54
  whitespace
1          0
2          0

── Variable type: numeric ──────────────────────────────────────────────
  skim_variable   n_missing complete_rate    mean    sd     p0
1 Year                    0             1 2003.    4.59 1995
2 Smoke everyday          0             1   16.6   3.98    3.6
3 Smoke some days         0             1    4.84  1.16    1.3
4 Former smoker           0             1   24.3   3.50    9.9
5 Never smoked            0             1   54.3   5.60   39.5
     p25    p50     p75   p100 hist
1 1999   2003   2007    2010  ▇▇▇▇▇
2   13.9   16.7   19.1    29.1 ▁▃▇▃▁
3    4.2    4.9    5.52    8.5 ▁▃▇▃▁
4   22.9   24.5   26.2    33.4 ▁▃▇▅▁
5   51.1   53.5   56.2    83.7 ▇▃▁▁▁
```

Table 2

(First check the relationship between year and never smoke, year and smoke everyday based states by applying ggplotly)

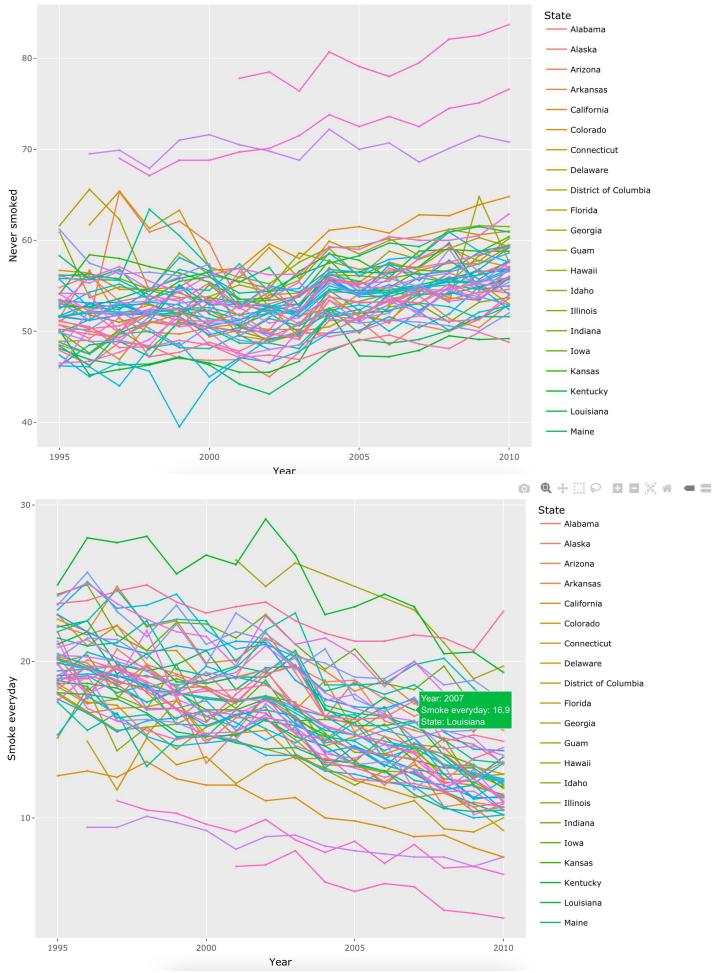


Table 3

(According to the data table, it shows that the average number of smokers in this data sets decreases as time goes by, and the average number of people who never smoke increases as time goes by)

Year	S_everyday_avg	S_someday_avg	S_former_avg	S_never_avg
1995	20.0137254901961	2.66470588235294	25.2274509803922	52.0764705882353
1996	19.6339622641509	4.1377358490566	23.8584905660377	52.3660377358491
1997	18.9555555555556	4.27222222222222	23.7759259259259	52.9814814814815
1998	18.4925925925926	4.56851851851852	24.0740740740741	52.8444444444444
1999	18.0611111111111	4.7962962962963	23.7092592592593	53.3925925925926
2000	17.5611111111111	5.06851851851852	24.4296296296296	52.9259259259259
2001	17.4053571428571	5.58035714285714	24.3160714285714	52.6714285714286
2002	18.0375	4.98214285714286	24.2732142857143	52.6857142857143
2003	17.1714285714286	5.13035714285714	24.6285714285714	53.0392857142857
2004	15.6481481481481	5.47592592592593	23.3444444444444	55.5481481481482

1–10 of 16 rows

	Year	S_everyday_avg	S_someday_avg	S_former_avg	S_never_avg
	2005	15.2290909090909	5.37454545454545	24.5072727272727	54.8672727272727
	2006	14.7909090909091	5.14363636363636	24.4945454545455	55.5236363636364
	2007	14.7428571428571	5.12857142857143	24.2589285714286	55.8303571428571
	2008	13.7714285714286	4.95535714285714	24.6964285714286	56.5446428571429
	2009	13.1464285714286	5.08392857142857	24.8267857142857	56.8875
	2010	12.8571428571429	4.85178571428571	24.7107142857143	57.5410714285714

11–16 of 16 rows

Previous 1 2 Next

Table 4:
(the relationship between the mean value of the number of 4 different smoking status and Year)

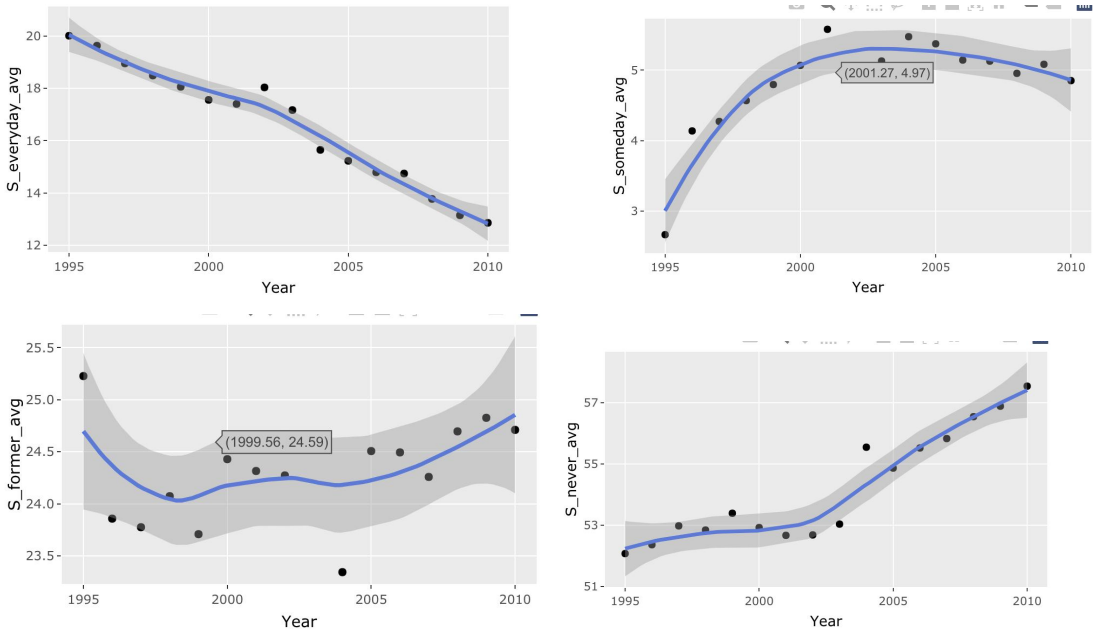
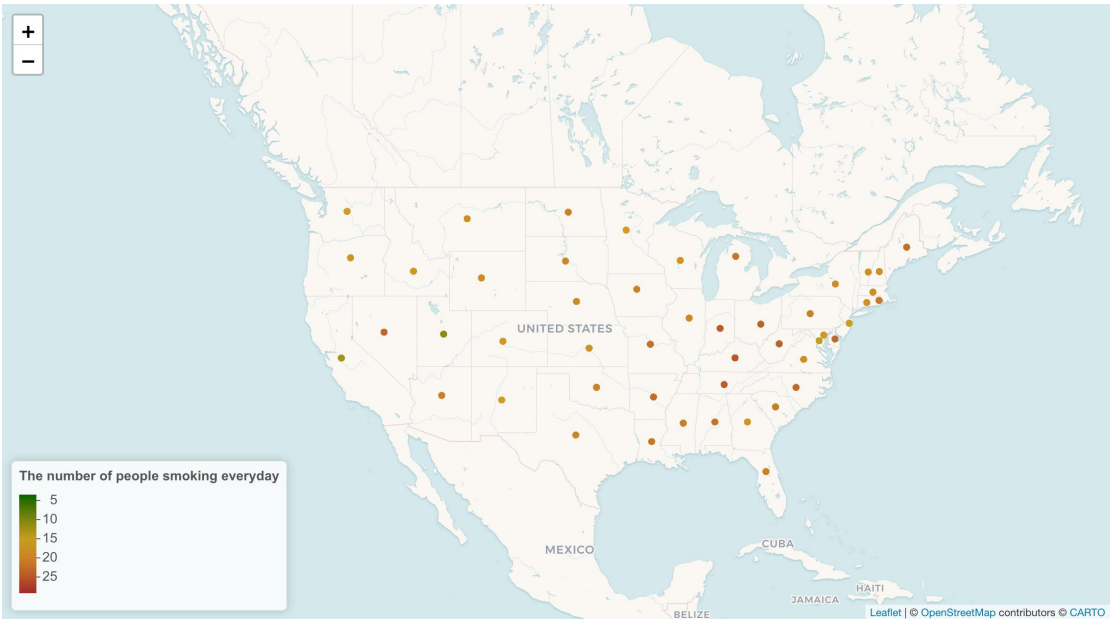
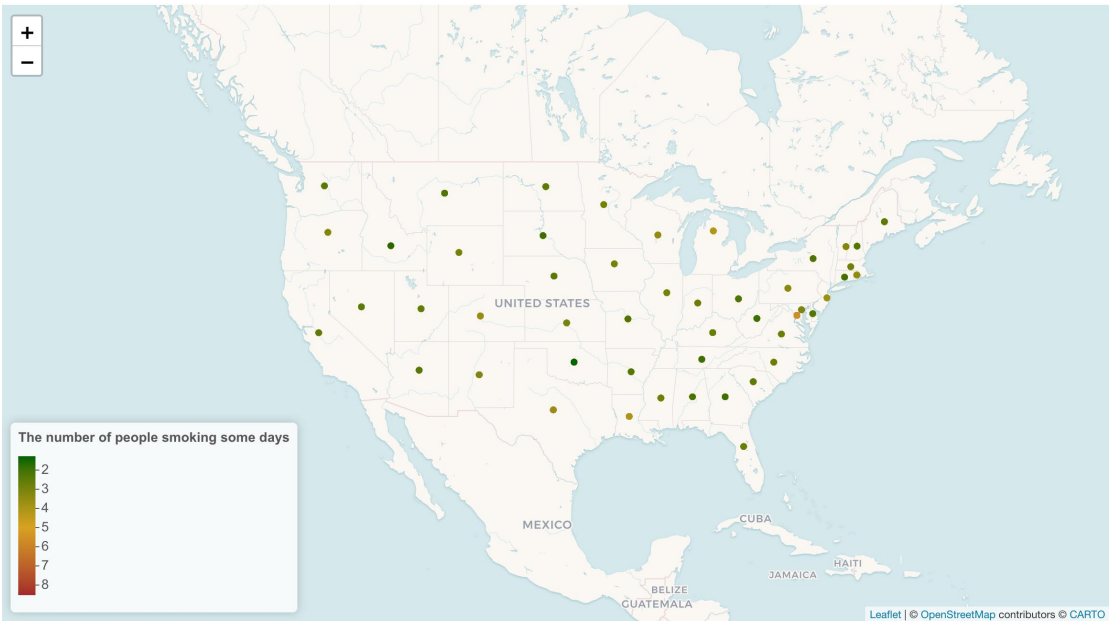
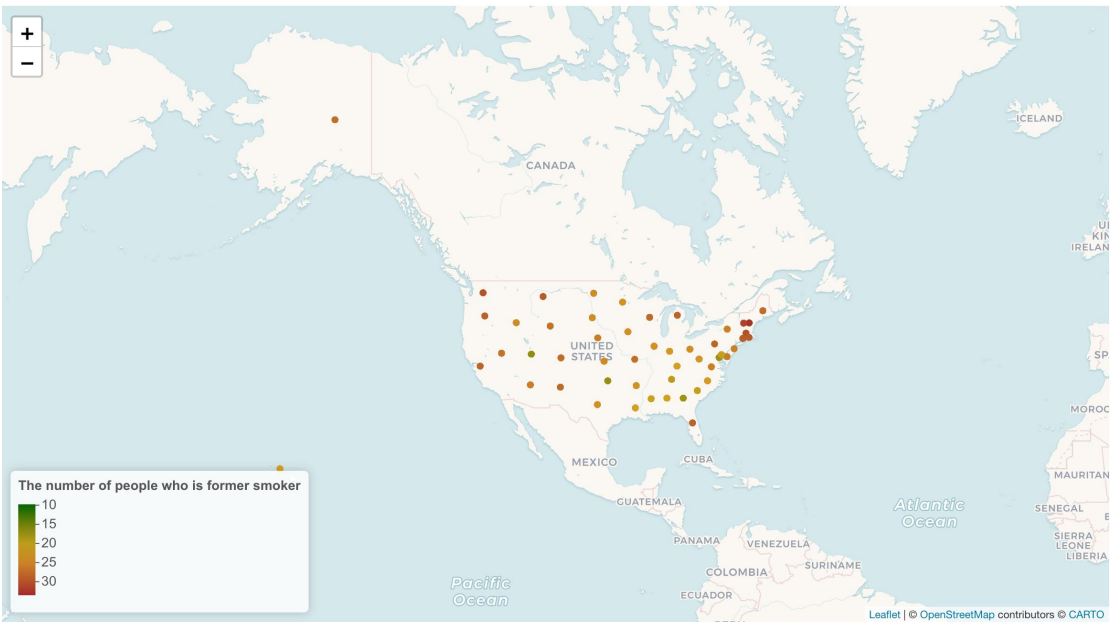
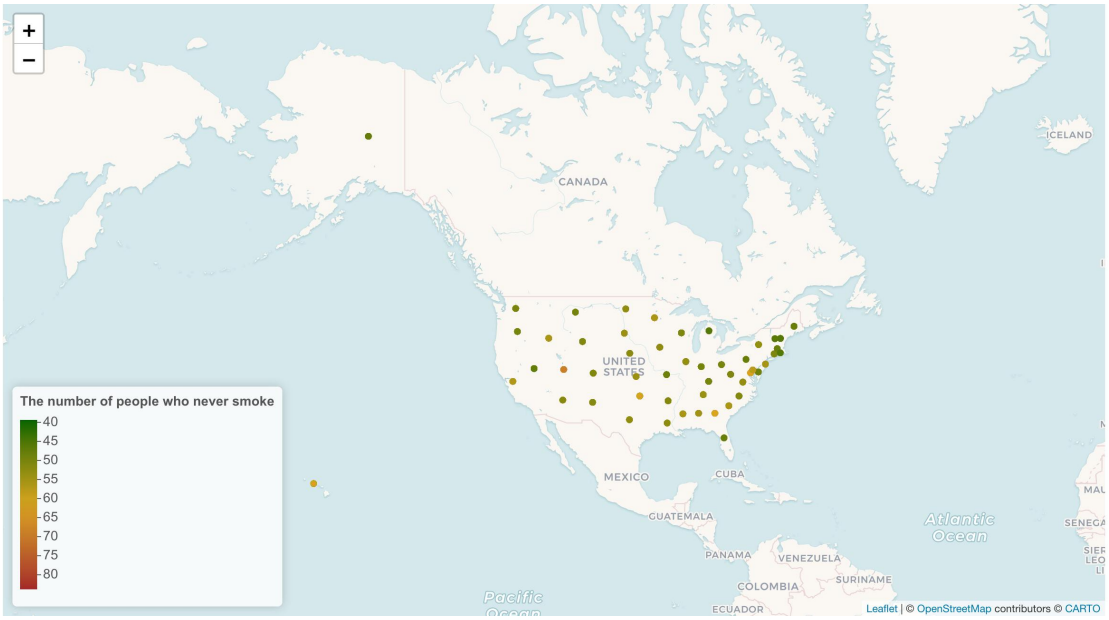


Table 5:
(the leaflet map showing the concentration of the number of smoking everyday, smoking someday, former smoker and never smoke on different state)


+
−
UNITED STATES
MEXICO
CUBA
HAITI
JAMAICA
BELIZE
The number of people smoking everyday
5
10
15
20
25
Leaflet | © OpenStreetMap contributors © CARTO


+
−
UNITED STATES
MEXICO
CUBA
HAITI
JAMAICA
BELIZE
GUATEMALA
The number of people smoking some days
2
3
4
5
6
7
8
Leaflet | © OpenStreetMap contributors © CARTO


+
−
ICELAND
CANADA
UNITED STATES
MEXICO
CUBA
GUATEMALA
PANAMA
VENEZUELA
COLOMBIA
SURINAME
ECUADOR
Atlantic Ocean
Pacific Ocean
SENEGAL
SIERRA LEONE
LIBERIA
The number of people who is former smoker
10
15
20
25
30
Leaflet | © OpenStreetMap contributors © CARTO


+
−
ICELAND
CANADA
UNITED STATES
MEXICO
CUBA
GUATEMALA
PANAMA
VENEZUELA
COLOMBIA
SURINAME
ECUADOR
Atlantic Ocean
Pacific Ocean
The number of people who never smoke
40
45
50
55
60
65
70
75
80
Leaflet | © OpenStreetMap contributors © CARTO