

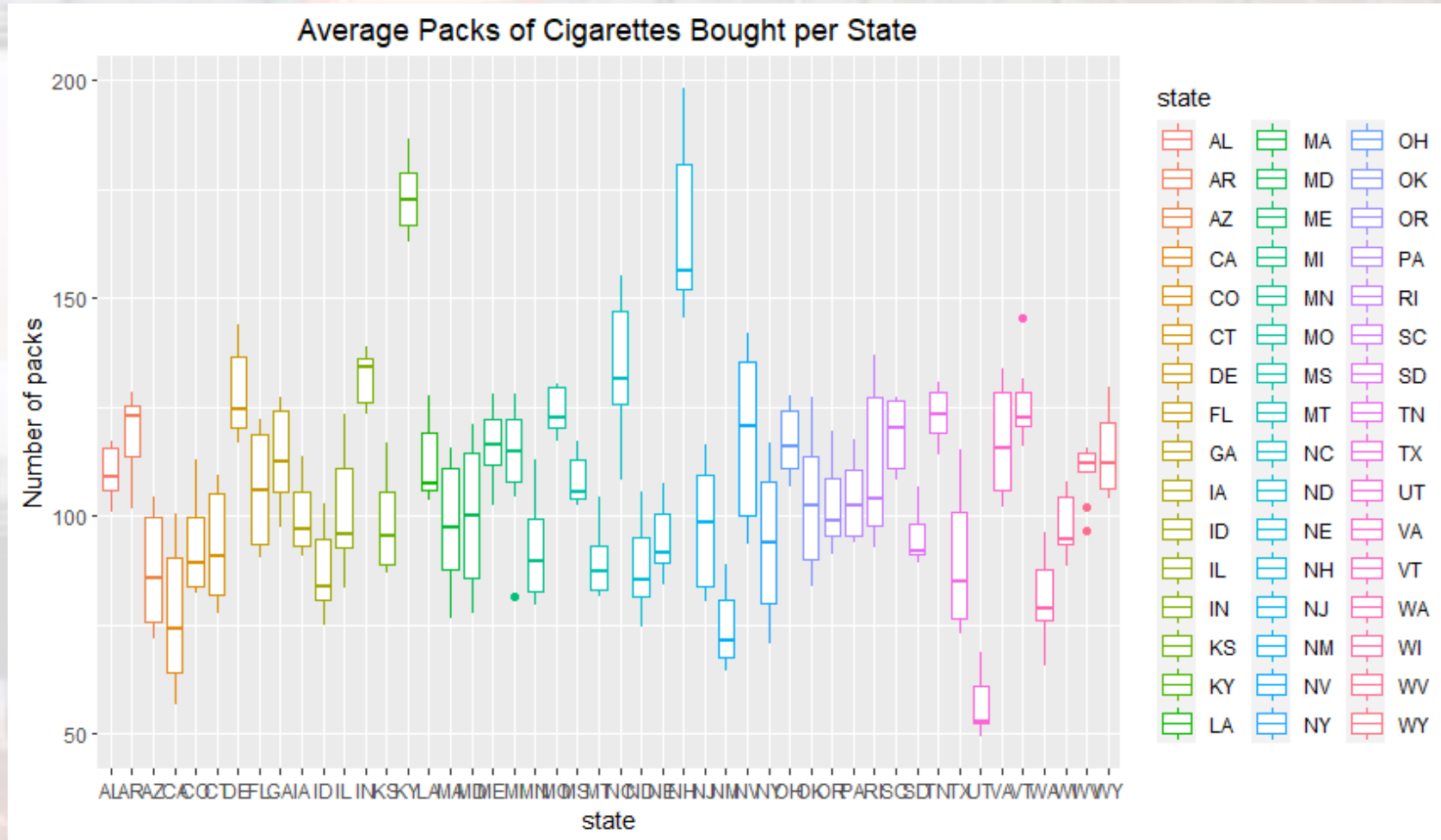
The background of the slide is a blurred image of cigarette packs. The top row features Camel packs with their iconic camel logo. Below them are Winston packs with a red and white design. The bottom row shows Pall Mall packs in various colors like red, blue, and green. A small price tag for \$5.75 is visible on the Camel packs.

A Data-Driven Analysis of Cigarette Consumption Patterns

By

Jackie Harmon

Average number of packs per capita by state



States with **highest** and **lowest** number of packs

Top 10	State	Mean
1	KY	174
2	NH	166
3	NC	135
4	IN	132
5	DE	128
6	VT	126
7	MO	124
8	TN	124
9	AR	119
10	SC	119

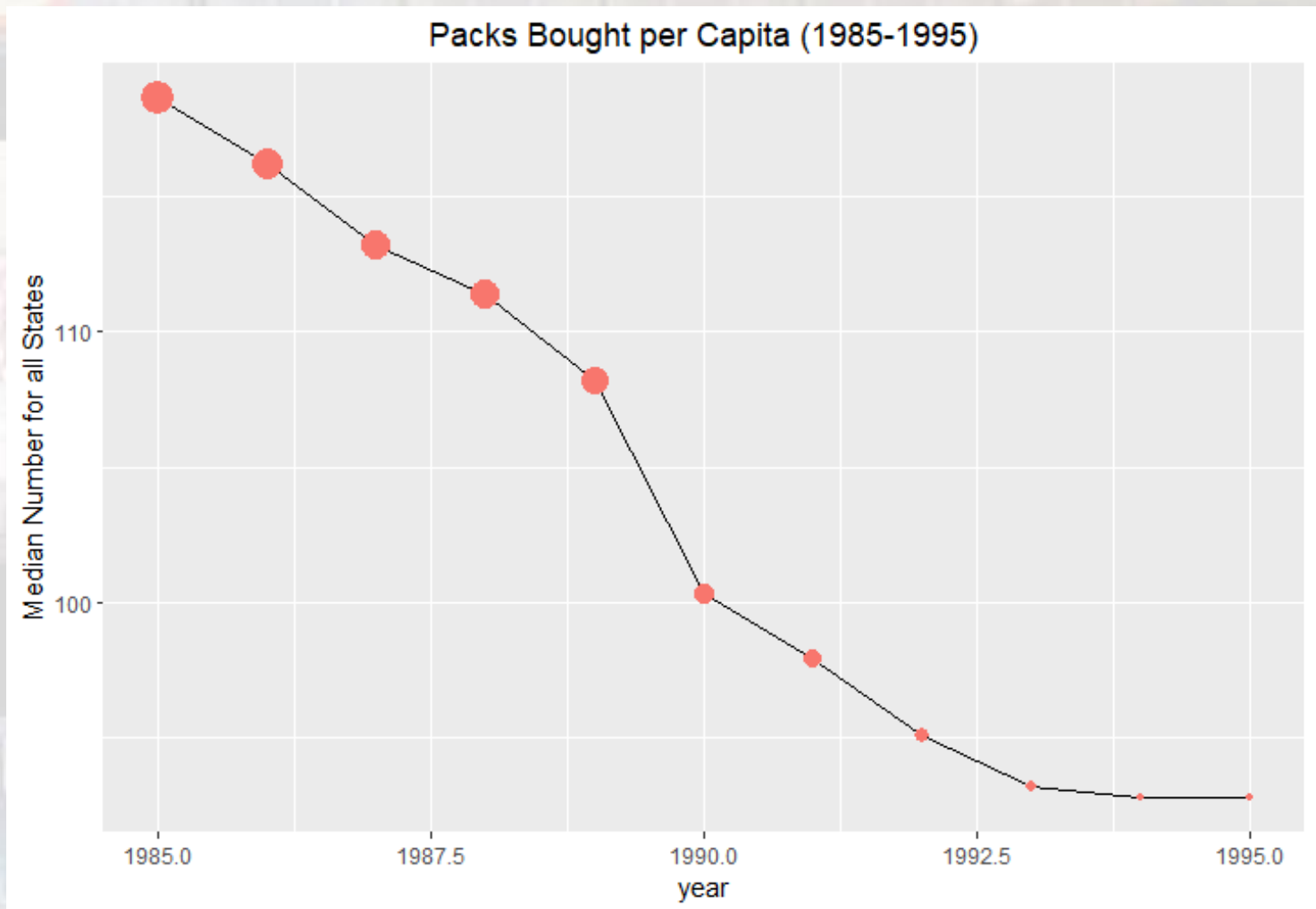
Bottom 10	State	Mean
50	UT	56.8
49	NM	74.4
48	CA	76.7
47	WA	81.0
46	ID	87.5
45	AZ	87.8
44	ND	88.4
43	MT	89.2
42	TX	89.8
41	MN	92.2

Median over all the states of the number of packs per capita for each year, plotted for the years from 1985 to 1995.

Overall, the cigarette usage is decreasing from 1985 to 1995.

Specifically, there is:

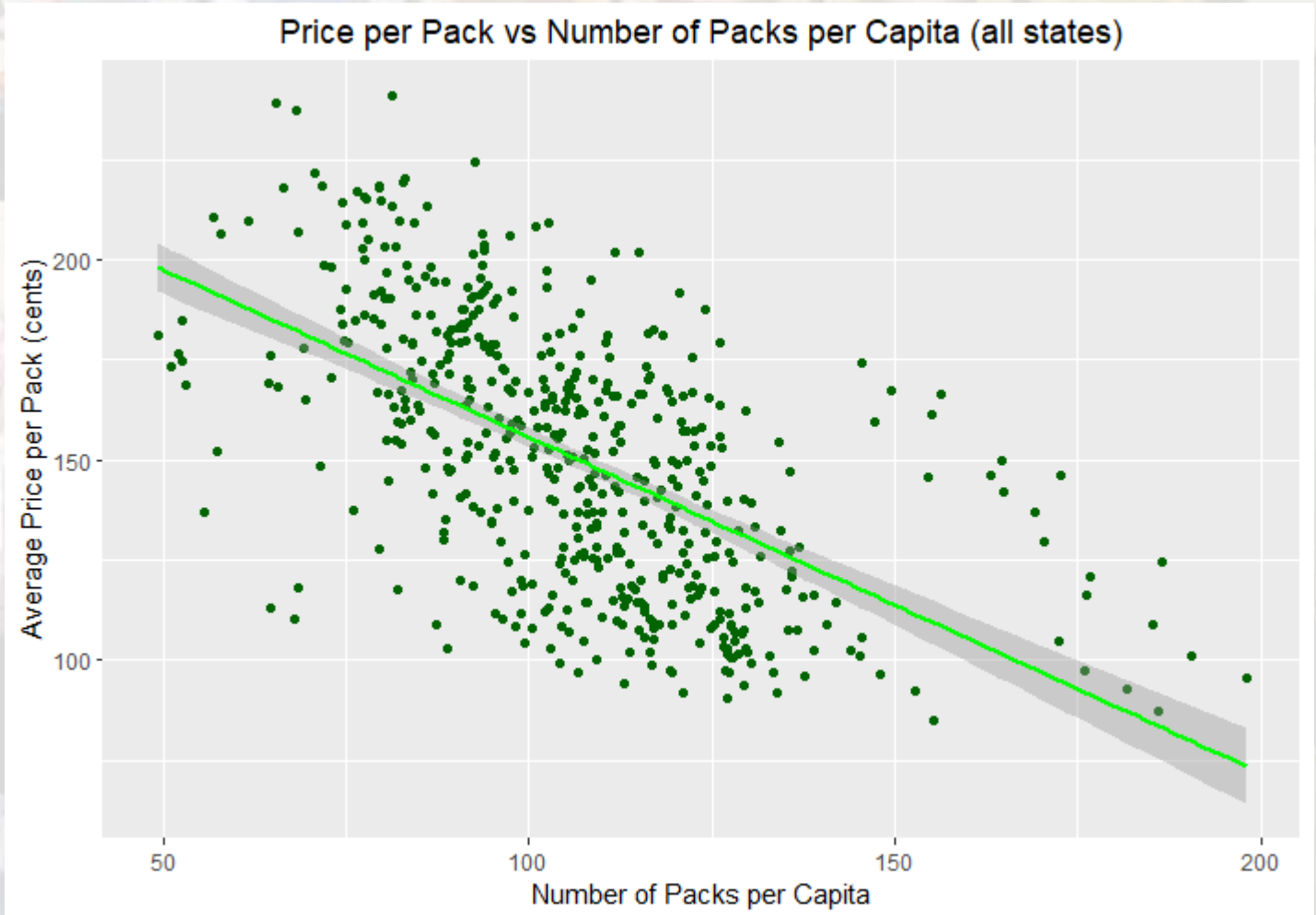
- a somewhat steady decrease in usage from 1985 to 1989
- a sharp decrease from 1989 to 1990
- a steady decrease from 1990 to 1993
- little to no decrease from 1993 to 1995 (leveling out)



Price per pack vs number of packs per capita for all states and years.

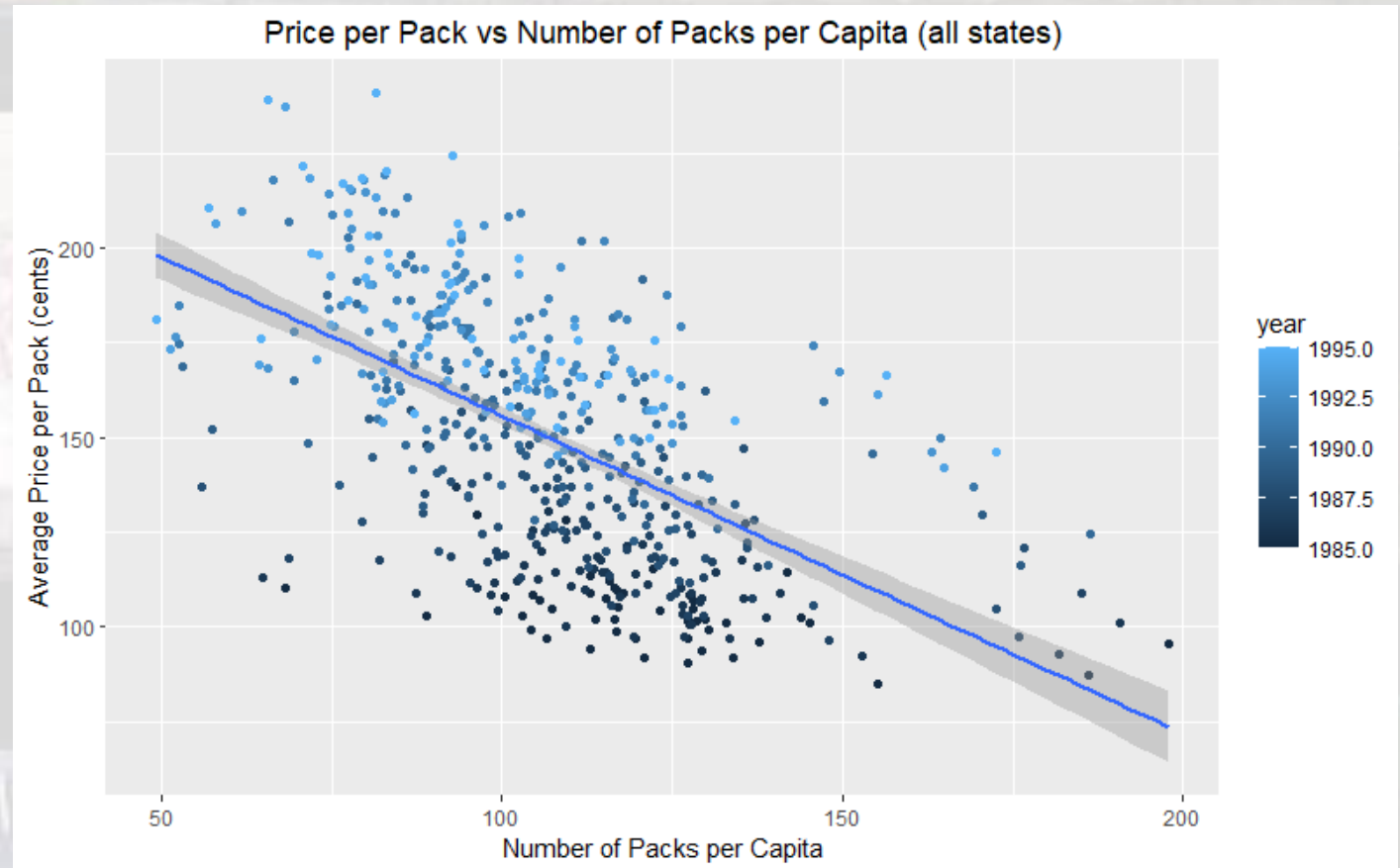
Price and per capita packs is negatively correlated, as shown by the green best fit line obviously decreasing from left to right. This is supported by the correlation coefficient being -0.585 , which suggests a negative relationship.

Negative correlation is expected because, in economics, the Law of Demand states that the quantity of a good demanded falls as the price rises, and vice versa.



Points for each year in a different color

- The darker data points are concentrated below the best fit line and to the right, meaning that in the years closer to 1985 the price of cigarette packs were lower and more of them were bought.
- The lighter data points are concentrated above the best fit line and to the left, meaning that in the years closer to 1995 the price of cigarette packs were higher and less of them were bought.
- The data is still negatively correlated.

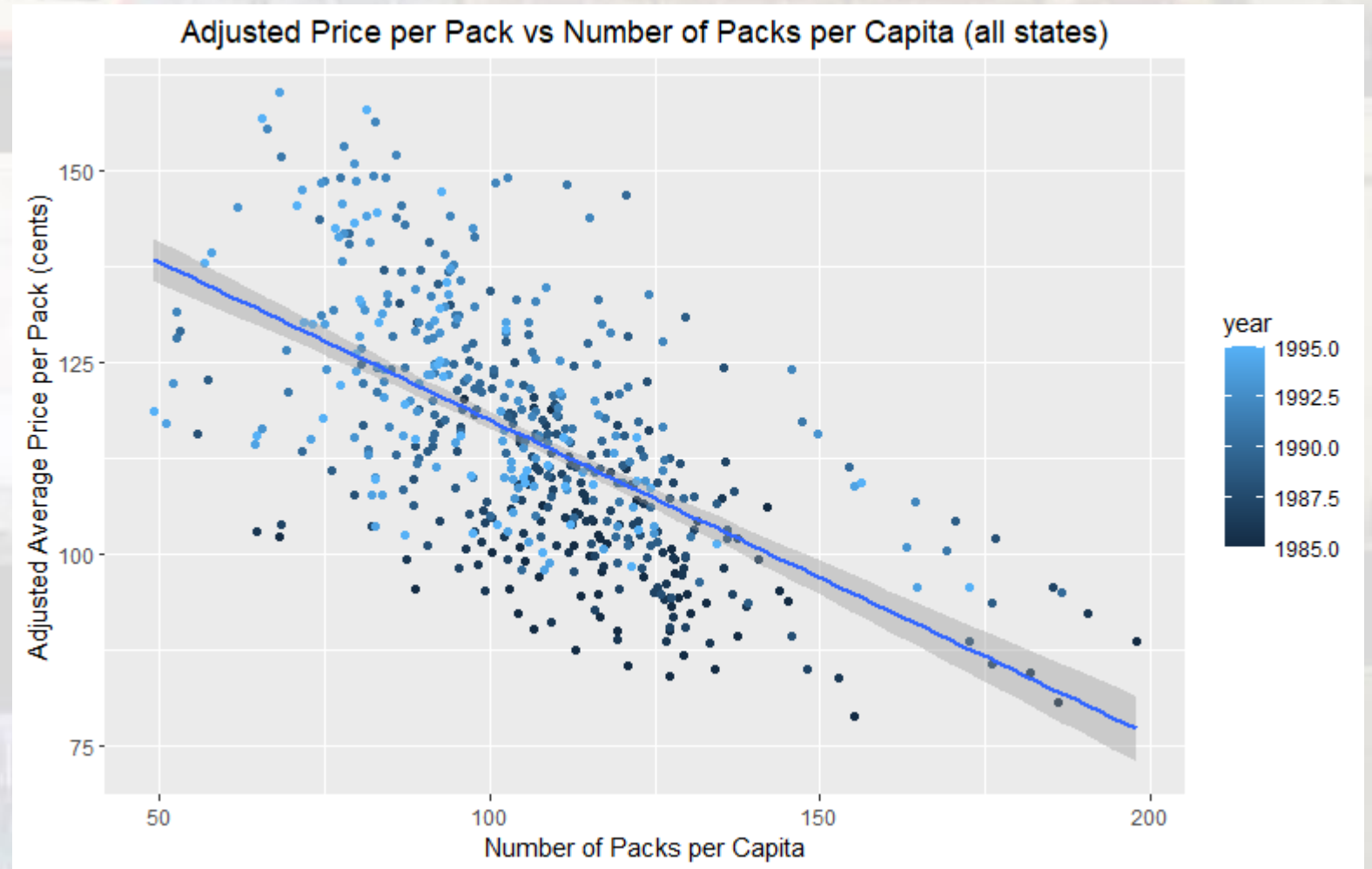


Linear regression

The model has a 0.3427 Multiple R-squared value, meaning that there is a 34.27% variability of average price that is explained by the number of packs bought.

Adjusted price for inflation

Adjusting for inflation changes the variability of the model to 37.69%. It also changes the distribution of the light dots, which now are frequently found below the best fit line as well as above. They are still concentrated to the left. This means that in the years closer to 1995, less packs of cigarettes were bought, regardless of price. The data is still negatively correlated.



Is the number of packs per capita in 1995 significantly different than the number of packs per capita in 1985?

Create a data frame with just the rows from 1985. Create a second data frame with just the rows from 1995. Then, from each of these data frames, get a vector of the number of packs per capita.

The p value is $2.2e-16$, which is lower than the alpha level of 0.05, so there is a statistically significant difference in number of packs per capita between 1995 and 1985.

Relationship between total state personal income and the number of packs of cigarettes bought per state

- Income and the number of pack of cigarettes are negatively correlated, because the best fit line obviously goes down as the graph moves from left to right. This is supported by the correlation coefficient being -0.255 .
- Most of the data points are on the lefthand side of the graph, suggesting more cigarettes are bought by people with lower incomes.
- The p value is $2.63e-9$, which is lower than the alpha level of 0.05 . Therefore, there is a statistically significant negative correlation between income and number of pack of cigarettes.

