Reed Ballesteros
MSDS-410-DL, Summer 2022
Dr. Mickelson
7/3/2022

Modeling Assignment #2: Fitting and Interpreting Simple Linear Regression Models

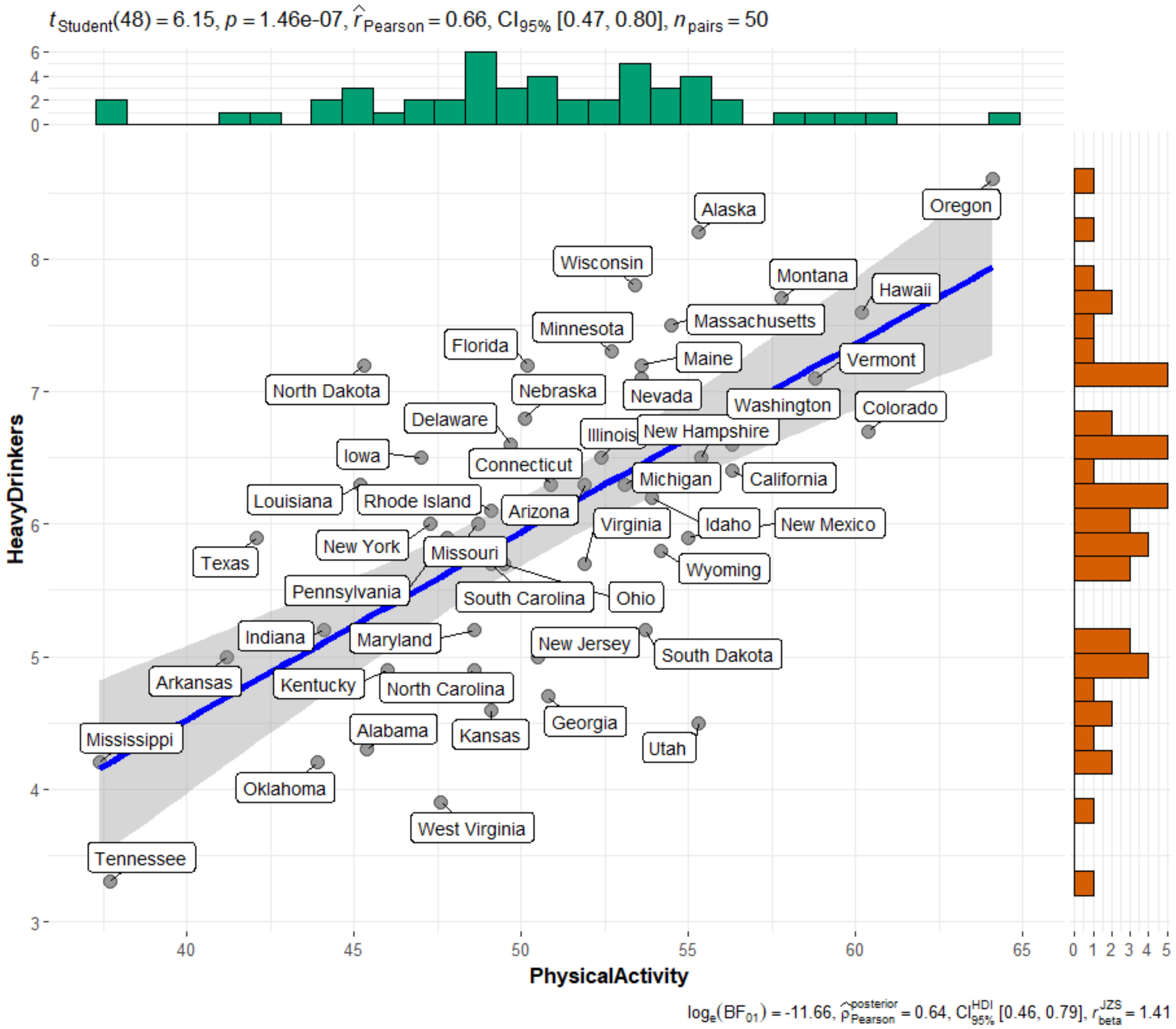
What stories can be told with data? In this exercise we will attempt to find a theme in the USStates census data, which tracks, by state, region, population (in millions), average household income (by thousands of US dollars), and state proportions (by percentage) of people based on the following properties:

- high school completion
- college completion
- smoking
- physical activity
- obesity
- non-white population
- heavy drinkers
- people raised in two-parent households
- insured

This report will discuss findings in the USStates data with some strong correlations found among states and between the properties above.

Physically Active People and Heavy Drinkers

States with higher rates of physically active people have a strong correlation with heavy drinking (**Pearson's r^2=0.66**). While the regression line has a fairly small slope (**Heavy Drinkers = -1.14 + 0.14 x Physical Activity**) we should take into account the range of values, particularly the heavy drinkers range between about 3.3% to just over 9% max.

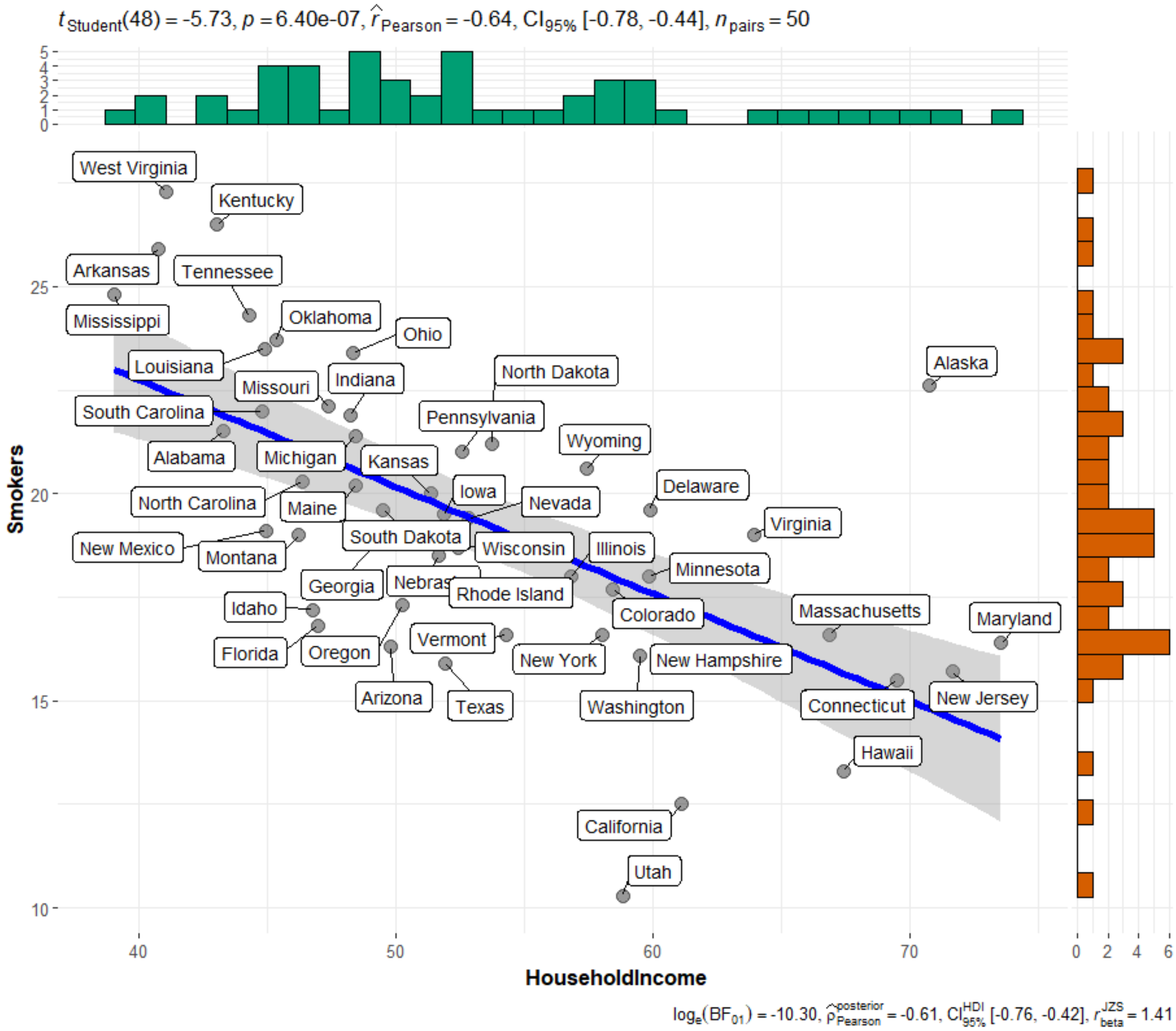


The heavy drinkers category also has the lowest percentage average (just over 6%) but relative to the rest of the US population we would like to see a generally low percentage of heavy drinkers. That being said, the regression equation infers those states with higher physical activity rates have higher heavy drinker rates, as with lower physical activity rates yield lower heavy drinking rates. This is an interesting observation as it would be perceived that higher physical activity rates would yield lower rates, but the data and this correlation shows the opposite. This is accentuated with the state of Oregon, a state known for its strong collegiate athletics from schools like the University of Oregon and Oregon State University and is also home to the headquarters of athletic apparel brands such as Nike, Adidas, and Under Armour. But Oregon is also known for its beer, particularly its carb-heavy IPAs, which could be a factor with the state having the highest heavy drinkers rate in the dataset.

An interesting outlier to the regression line is Utah, which could play more into the perception of having an above average physical activity rate and a low heavy drinker rate. But the low heavy drinker rate in Utah could be due to the state having the most restrictive regulations in the US with the selling and purchasing of alcohol in general.

Household Income and People Who Smoke

The regression line between state average household income and smokers percentage (**Smokers = 33.09 – 0.259 x Household Income**) show a fairly strong negative correlation (**Pearson's r^2=-0.64**).

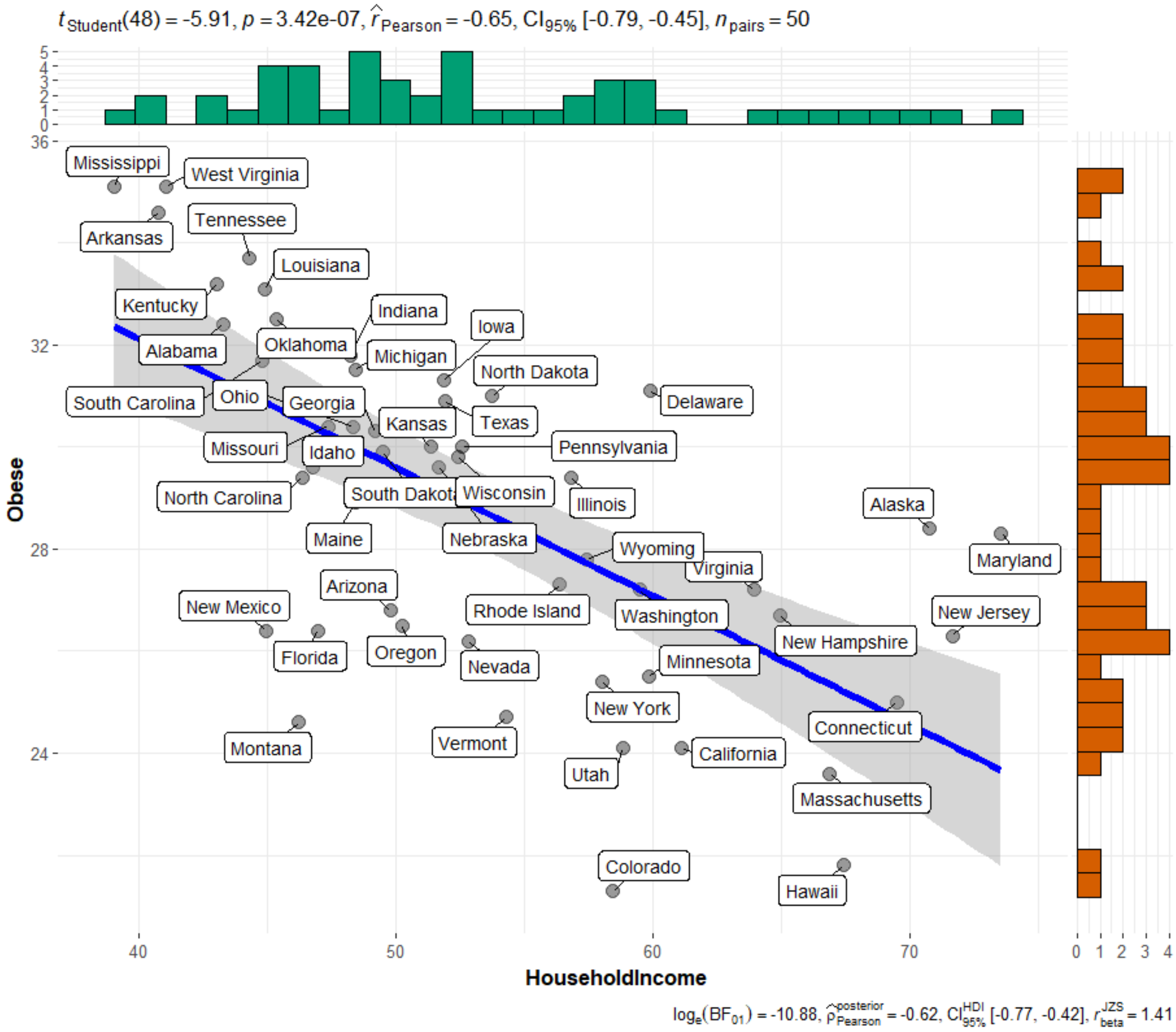


Poorer states show a higher rate of smokers while more affluent states have a lower percentage of smokers. Despite the cost of cigarettes today and its health affects, a theme we will be seeing in this report is that poorer states tend to be less healthy overall. There could be a possibility that poorer states have higher rates of smokers due to the stress of living in poverty and other ways to manage mental health is not affordable.

Alaska is an interesting outlier from the dataset in that it has both higher smoking rates and household income. While Alaska's higher income compared to most states is due to the higher cost of living. Alaska is also known to have one of the highest crime rates in the country. The high smoking rate could be due to dealing with the stress of both potential factors.

Household Income and Obesity

The regression line between household income and obesity (**Obesity = 42.176 - 0.25 x Household Income**) shows a fairly strong negative correlation (**Pearson's r^2=-0.65**).



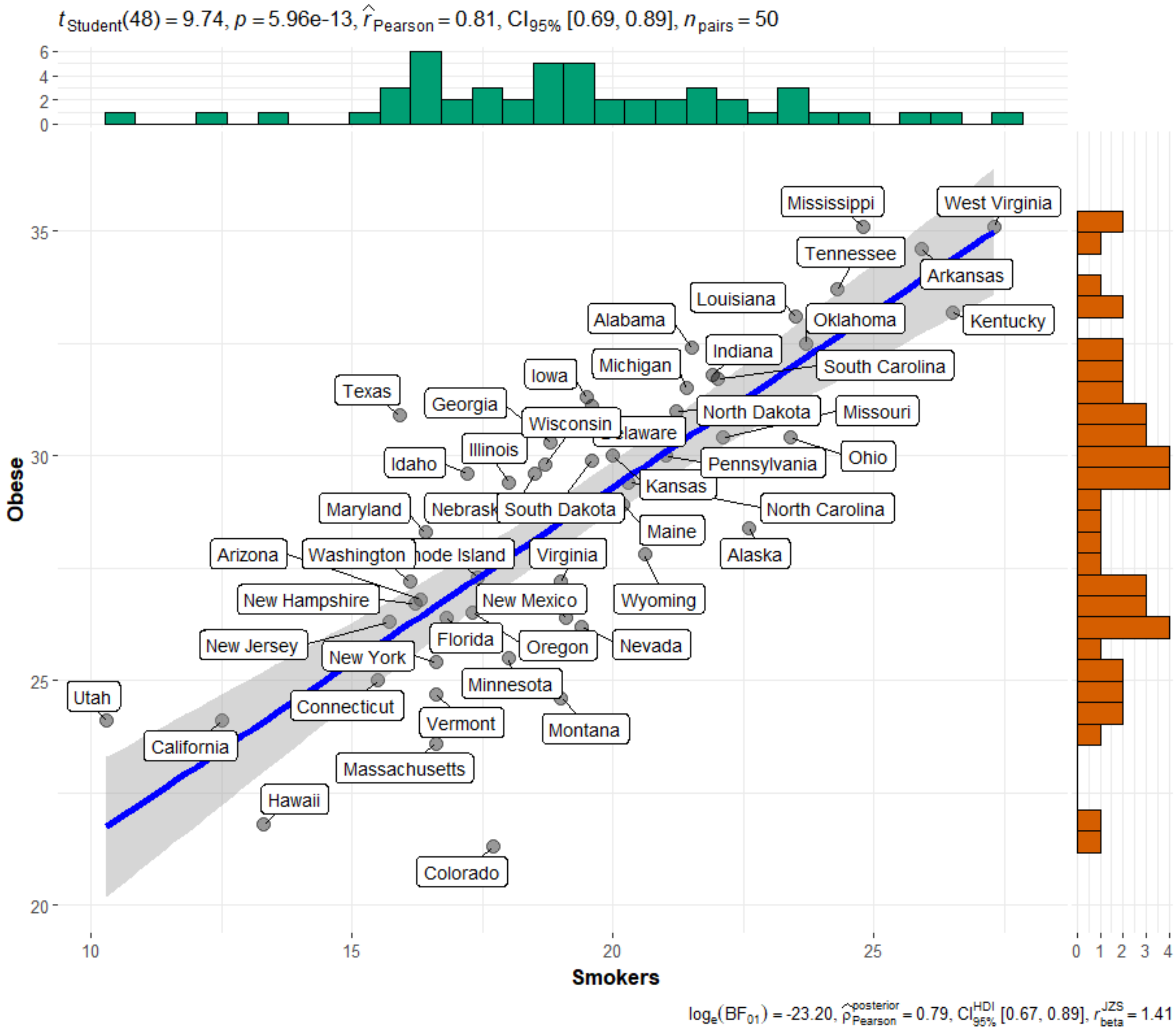
$t_{\text{Student}}(48) = -5.91$, $p = 3.42\text{e-}07$, $\hat{r}_{\text{Pearson}} = -0.65$, $\text{CI}_{95\%}$ [-0.79, -0.45], $n_{\text{pairs}} = 50$

$\log_e(\text{BF}_{01}) = -10.88$, $\hat{\rho}^{\text{posterior}}_{\text{Pearson}} = -0.62$, $\text{CI}^{\text{HDI}}_{95\%}$ [-0.77, -0.42], $r^{\text{JZS}}_{\text{beta}} = 1.41$

Heathy food choices are not cheap and may take time to prepare, which in turn states with lower average household incomes might have more of a dependency on fast-food, and food containing preservatives for higher shelf life and affordability. With these kinds of food options available to them, poorer states could have higher obesity rates.

Smokers and Obesity

The data shows a very strong correlation (**Pearson's r^2=0.81**) between states with a higher percentage of smokers and obesity:

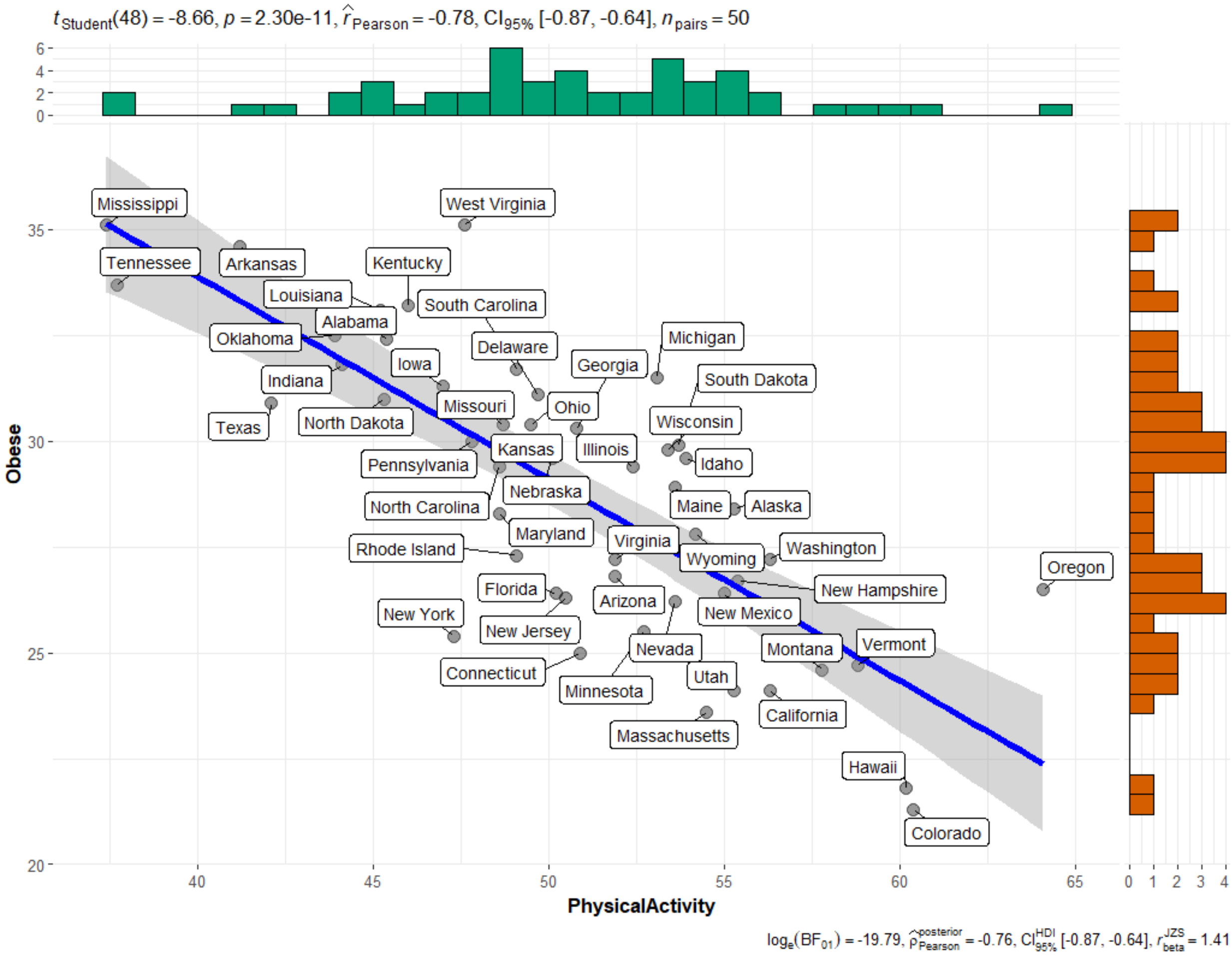


This is an interesting observation as the regression line (**Obesity = 13.71 + 0.48 x Smokers**) shows the opposite perception of smoking being used as a form of dieting. That being said, this particular correlation can be attributed their respective strong negative correlations to household incomes and physical activity. This graph helps support that poorer states have higher obesity and smoking rates, which can also show that poorer states are less healthy overall as well.

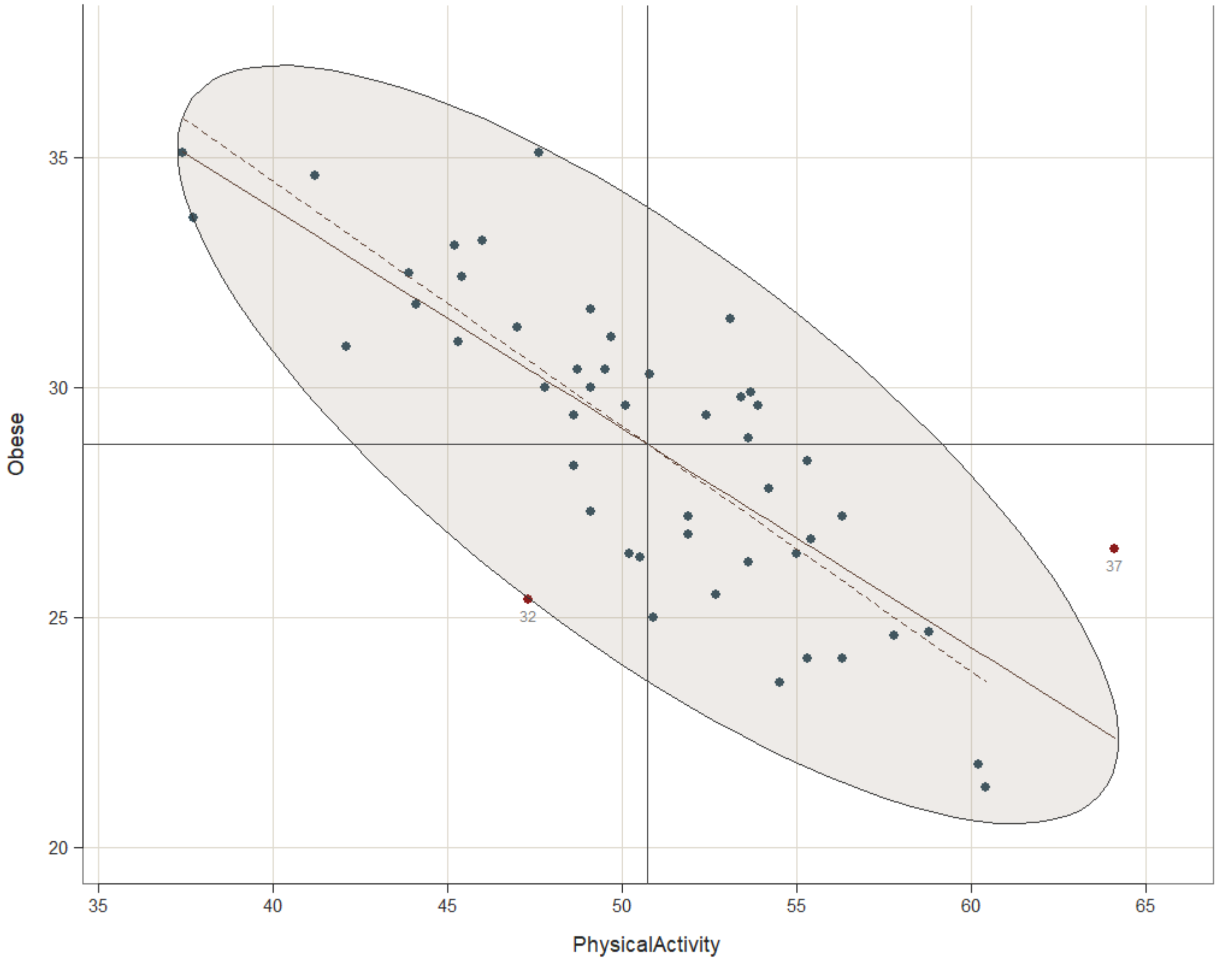
An interesting observation among the outliers of this correlation is Colorado. The state can be perceived as one of the 'healthier' states with the lowest obesity rate and the higher rates of physical activity. The state's high elevation makes it a destination for many athletes to train in and for people to have very active lifestyle, particularly skiing during the winter season. That being said, its smoking rate (17.7%) is just below average among all states (19.32%). Colorado is one of the first states to legalize marijuana for recreational use and the smoking rate could possibly include people that partake in that.

Physically Active People and Obesity

The USStates data shows a very strong negative correlation (**Pearson's r^2=-0.78**) between states' percentage of people that are physically active and their percentage of obesity:



From the linear regression line (**Obesity = 52.99 – 0.48 x Physical Activity**) we can infer those states with higher proportions of people that are physically active have lower obesity rates. In other words, populations with more active, healthier lifestyles have less obesity. An ellipse of the surrounding data shows a long and tight grouping of data:



An interesting outlier from the correlation, Oregon, has the highest percentage of the population physically active, but is also just below the national rate of obesity (28.8% national vs. 26.5% Oregon). As mentioned earlier, Oregon's physically active population could be due to its strong collegiate athletics and is the home of the headquarters to several major athletic apparel brands. Its largest city, Portland, has a very diverse community which also offer a wider variety of foods and help the state rank as one of the higher 'foodie' destinations in the nation (DeNike, 2022; McCann 2021). While the people of Oregon have a very active lifestyle, it seems they also like to eat and drink as well.

Summary

While this report did not utilize all of the properties in the USStates data, the most compelling story found in the data used could be the overall correlation of health to household income. States with higher income generally seem to be healthier in terms of lower rates of smoking and obesity while also having a higher percentage of their population participate more into physical activity. Higher incomes give people more options, opportunities, and choices for healthier living. People in poorer states might have to work more to earn a living to pay bills and simply get by, and by doing so they could be very limited in their choices for eating better and having a more active and healthier lifestyle.

The Analysis of Variance (ANOVA) between each of the strongly correlations above show very low p-values along with fairly large F-values to show that the correlations are not by chance but show a solid relationship between two given properties for each state.

References

DeNike, Max, 2022. “U.S. States Ranked by Their Food.” Far & Wide, January 20,2022. https://www.farandwide.com/s/us-states-ranked-food-2f86e72cb89c4904

McCann, Adam. 2021. “Best Foodie Cities in America”. WalletHub, October 5, 2021. https://wallethub.com/edu/best-foodie-cities/7522