S670 Final Project

The Fantastic Four
Erik "Human Torch" Parker
Emily "Invisible Woman" Rudman
Vinay "The Thing" Vernekar
Jervis "Mister Fantastic" Wang

Description of Data

Data Sources - Two primary data sources are:

a) Catalog.data.gov - The data contains details like store/restaurant proximity, food prices, community characteristics etc.

b) Factfinder.census.gov - Income details (median income) state and county level

PCT_OBESE_ADULTS10 - The percentage of obese adults in a county in 2010.

NATAMEN - The Natural Amenities scale, is an index developed by the USDA in 1999 which ranges from 1 to 7 and measures the desirability of a particular location (here a county) based on natural factors.

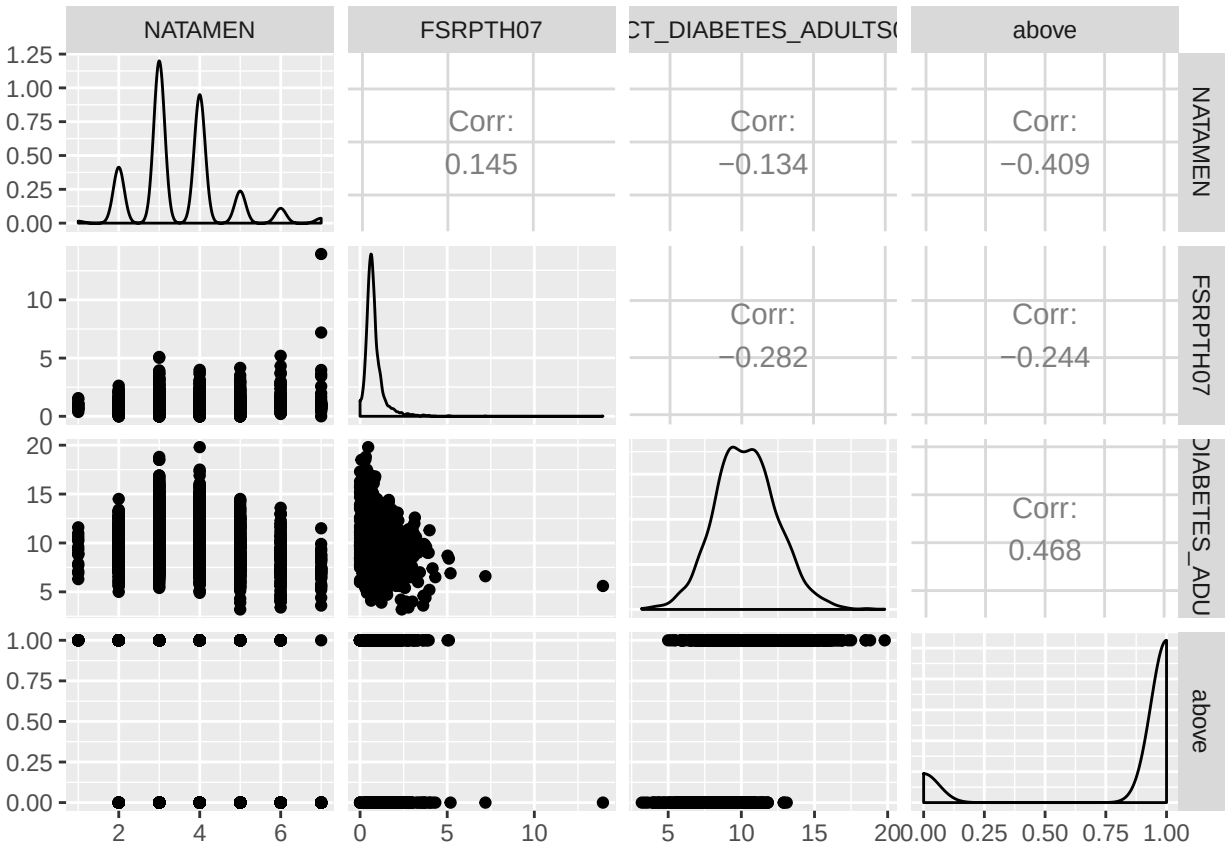
CONVSPTH07 - Is a measure of the number of convenience stores in a county, per 1000 residents, in 2007.

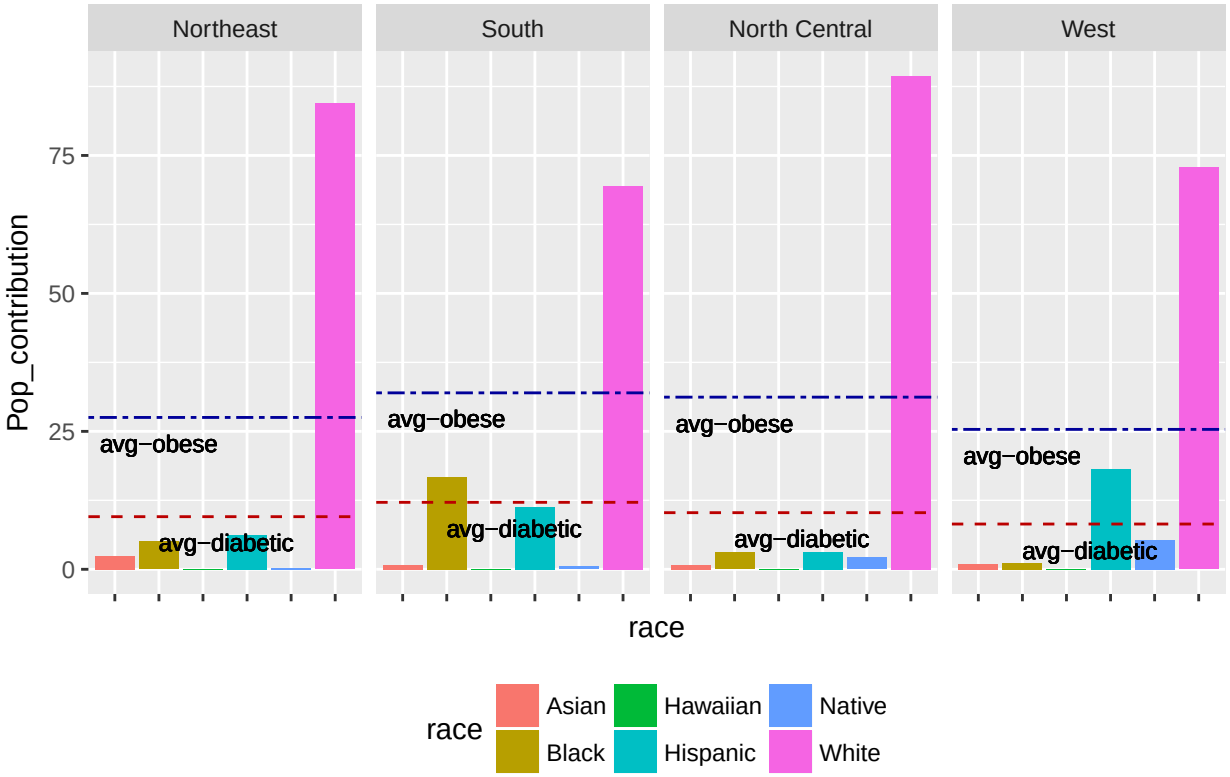
FSRPTH07 - Is the same measure but for full service restaurants in a county, per 1000 residents, in 2007.

PCT_DIABETES_ADULTS10 - The percentage of adults that have diabetes in a county in 2010.

Above - A binary variable that assigns author: - The Fantastic Four - Erik "Human Torch" Parker - Emily "Invisible Woman" Rudman - Vinay "The Thing" Vernekar - Jervis "Mister Fantastic" Wanga 1 to a county who is above the national average obesity rate of 26.7% and 0 otherwise.

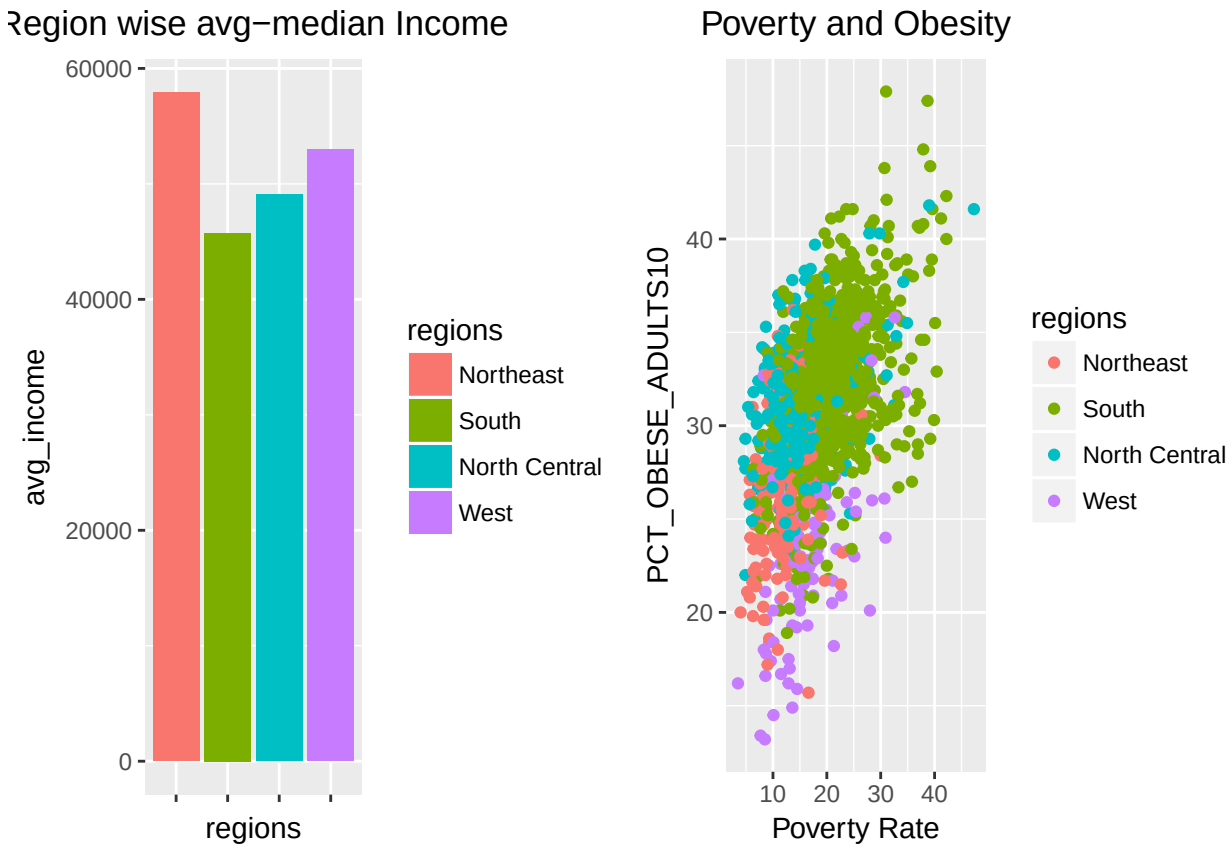
Poverty Rate - The poverty rate is the ratio of the number of people whose income falls below the poverty line.





Observations

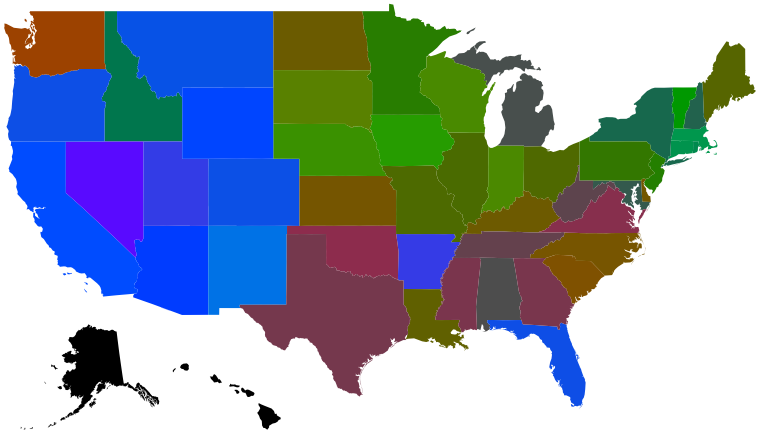
1) Regions have considerable difference in the race diversity
2) South and West regions have more diversity
3) West region has more obese rate and diabetes rate



Observations

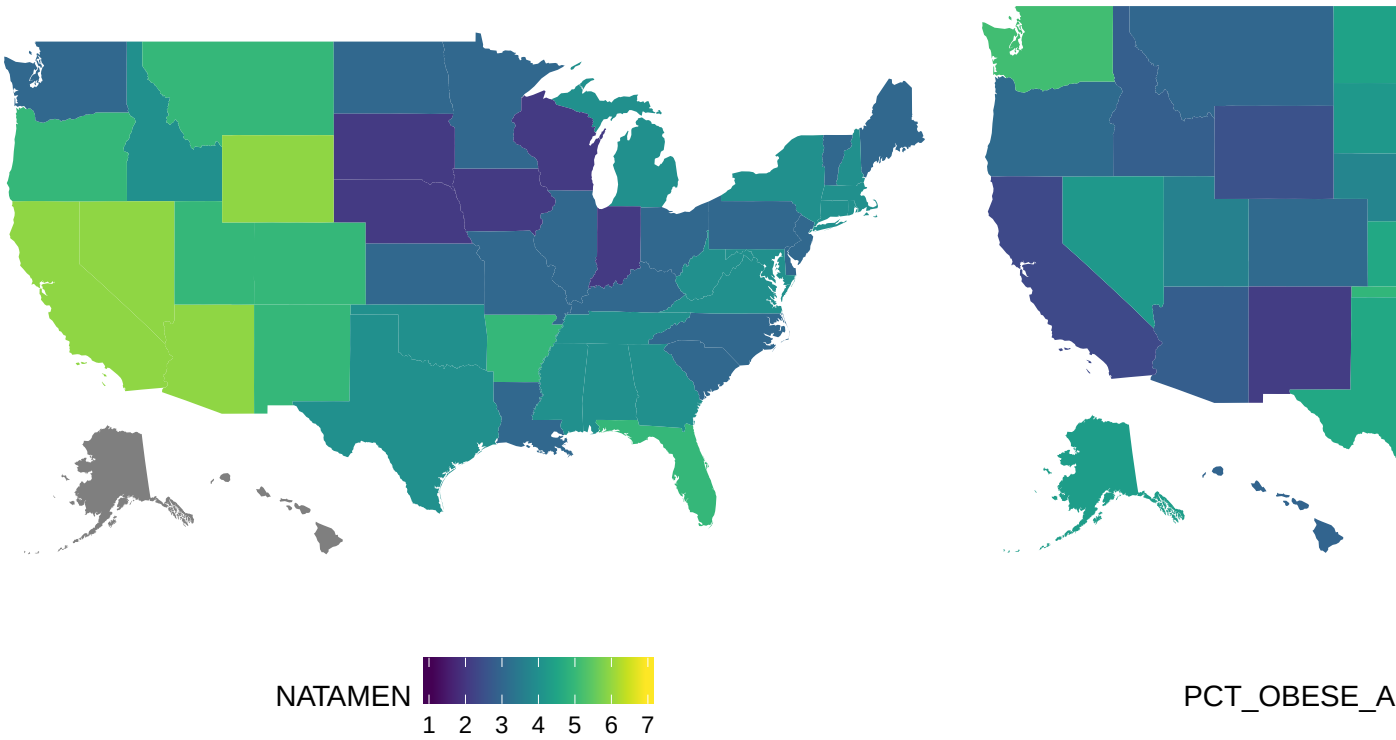
1) South has the lowest avg median income level
2) Earlier graphs indicate high obesity and diabetes rate in south
3) The scatter chart clearly indicates some relationship between Poverty and obesity

A heat map of Obesity Rate vs Natural Amenities Scale



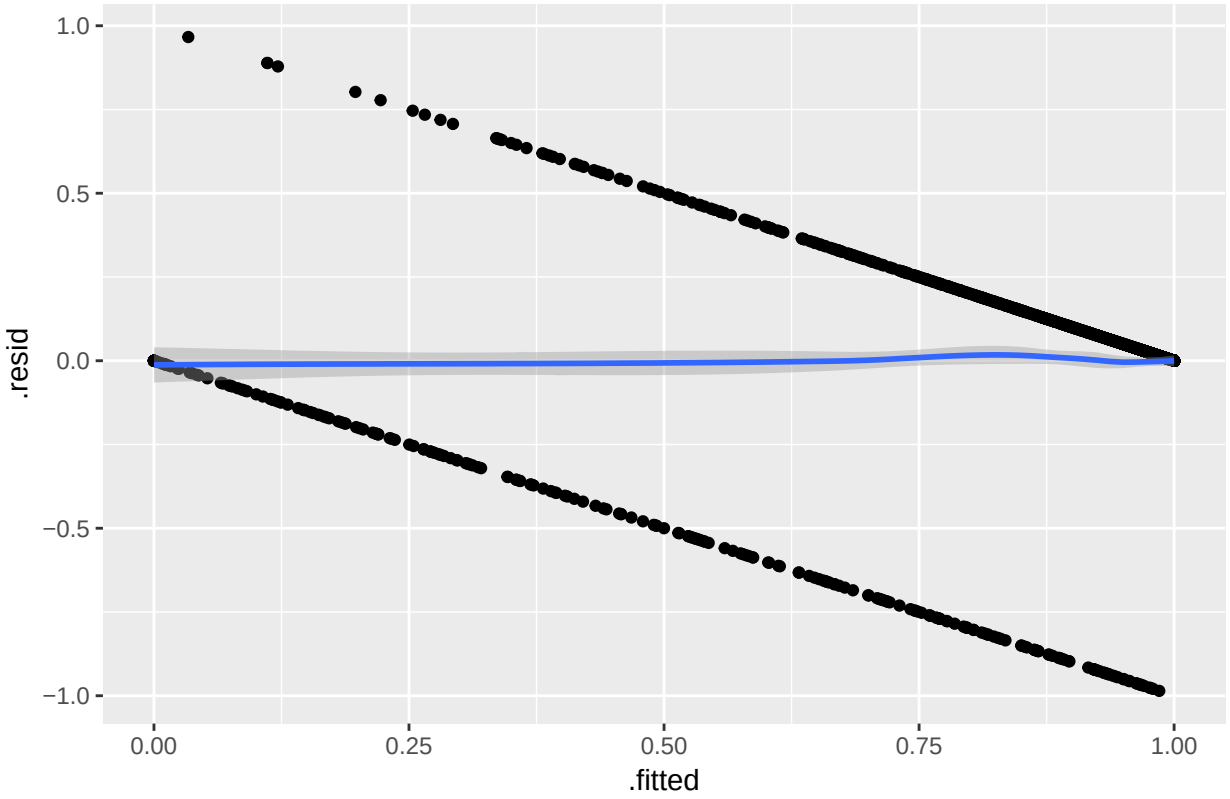
A heat map of Natural Amenities Scale averaged by states

A heat map of Obesity Rate

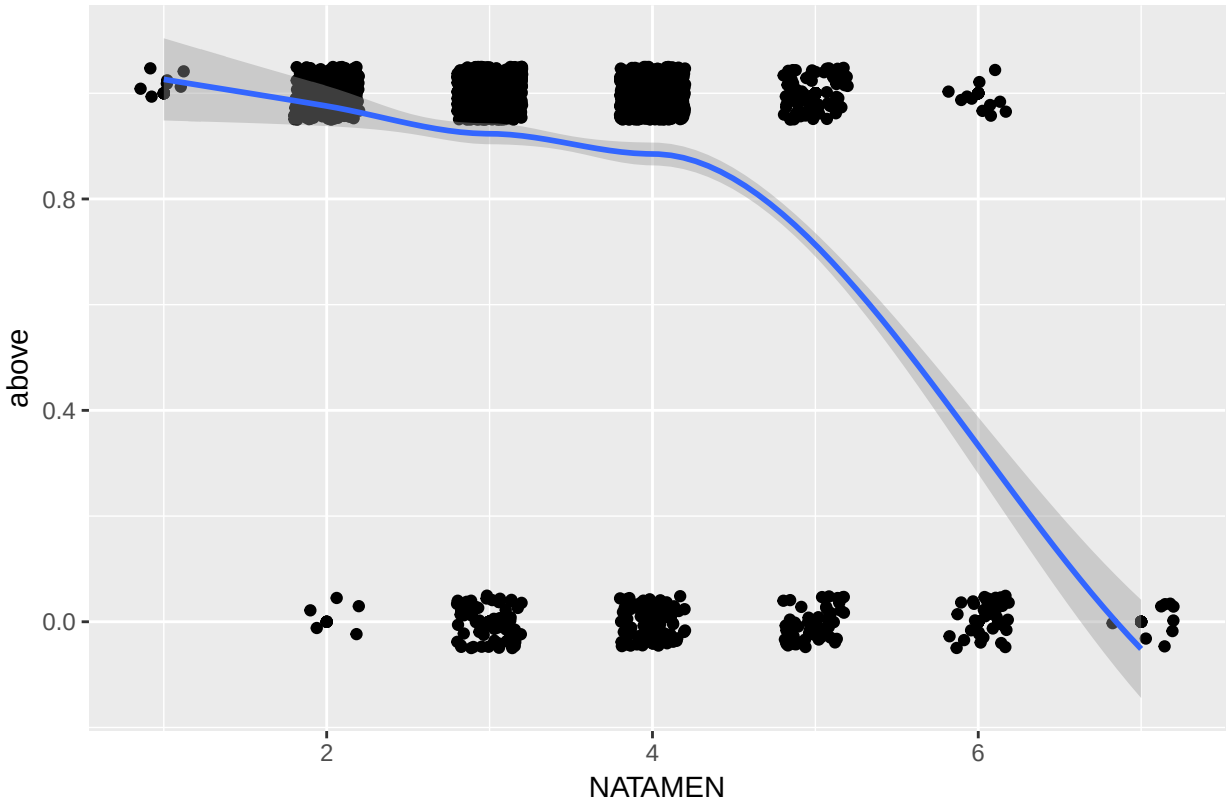


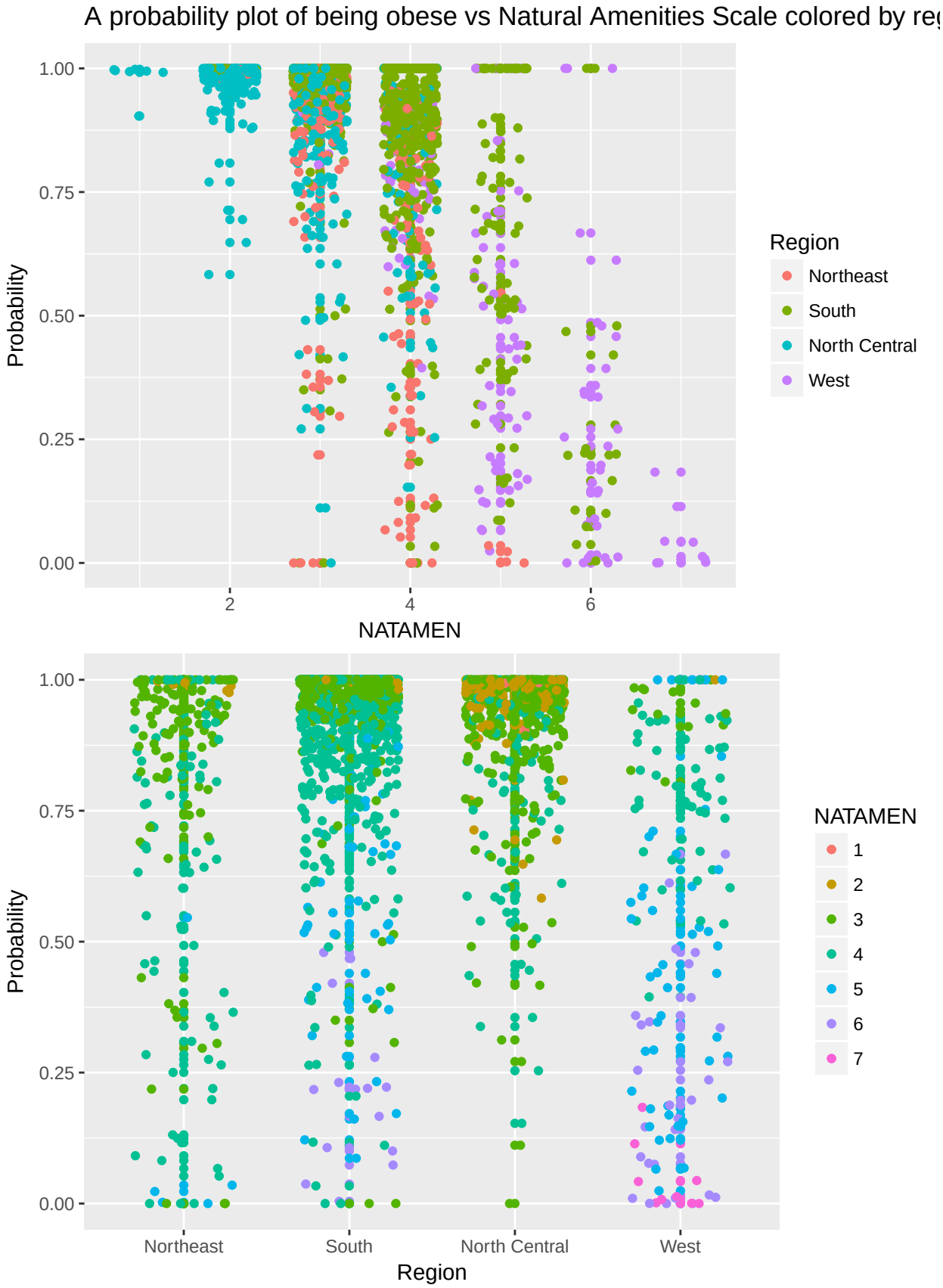

```
model.ob <- glm(above ~ FSRPTH07 + NATAMEN + POVRATE10, family = "binomial",
    data = model.data)

# summary(model.ob) Binomial model predicting if counties are above average
# obesity from convenience stores/1000 people, full service resturants/1000,
# and NATAMEN.  Residual deviance: reduces when poverty rate is included,
# reduces by ~200 pts AIC is also lower when we have poverty rate ( Better
# model)
```

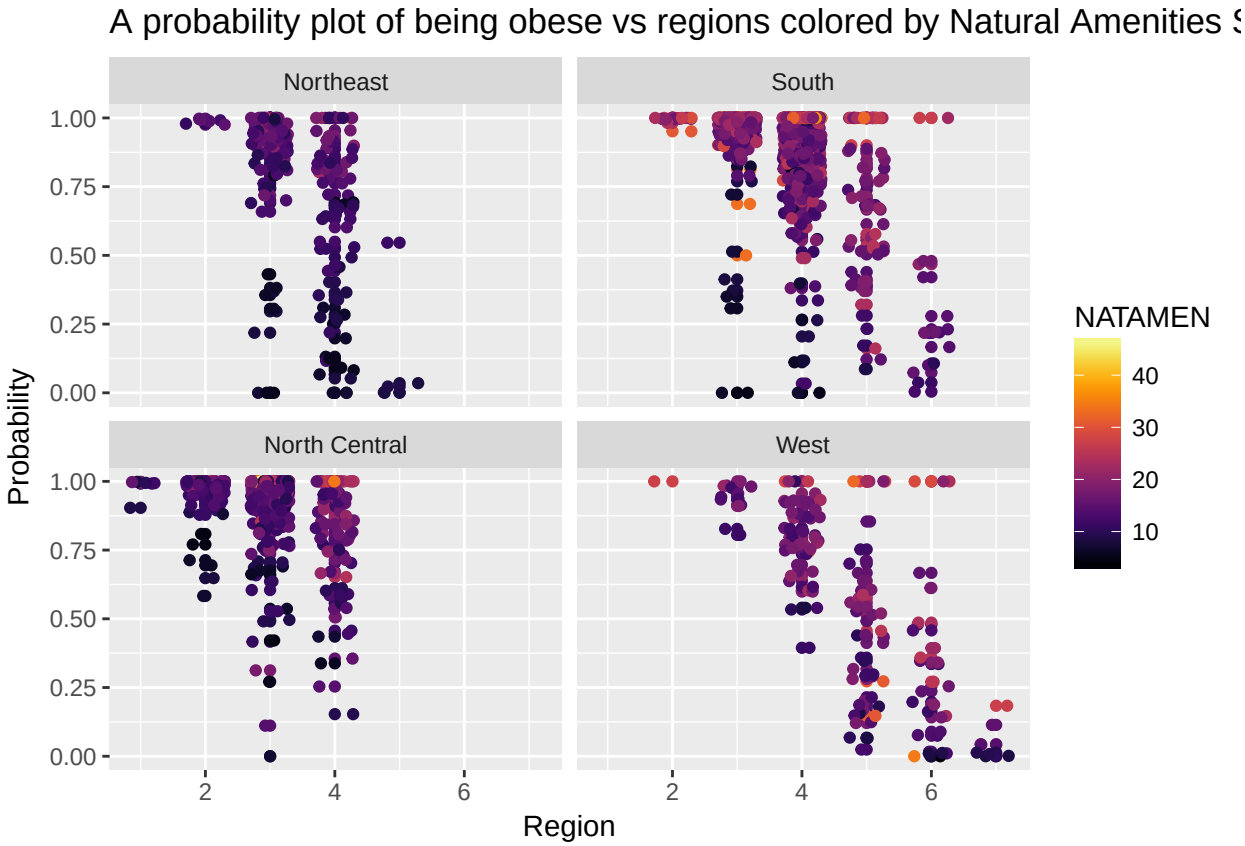
A residual plot of the model

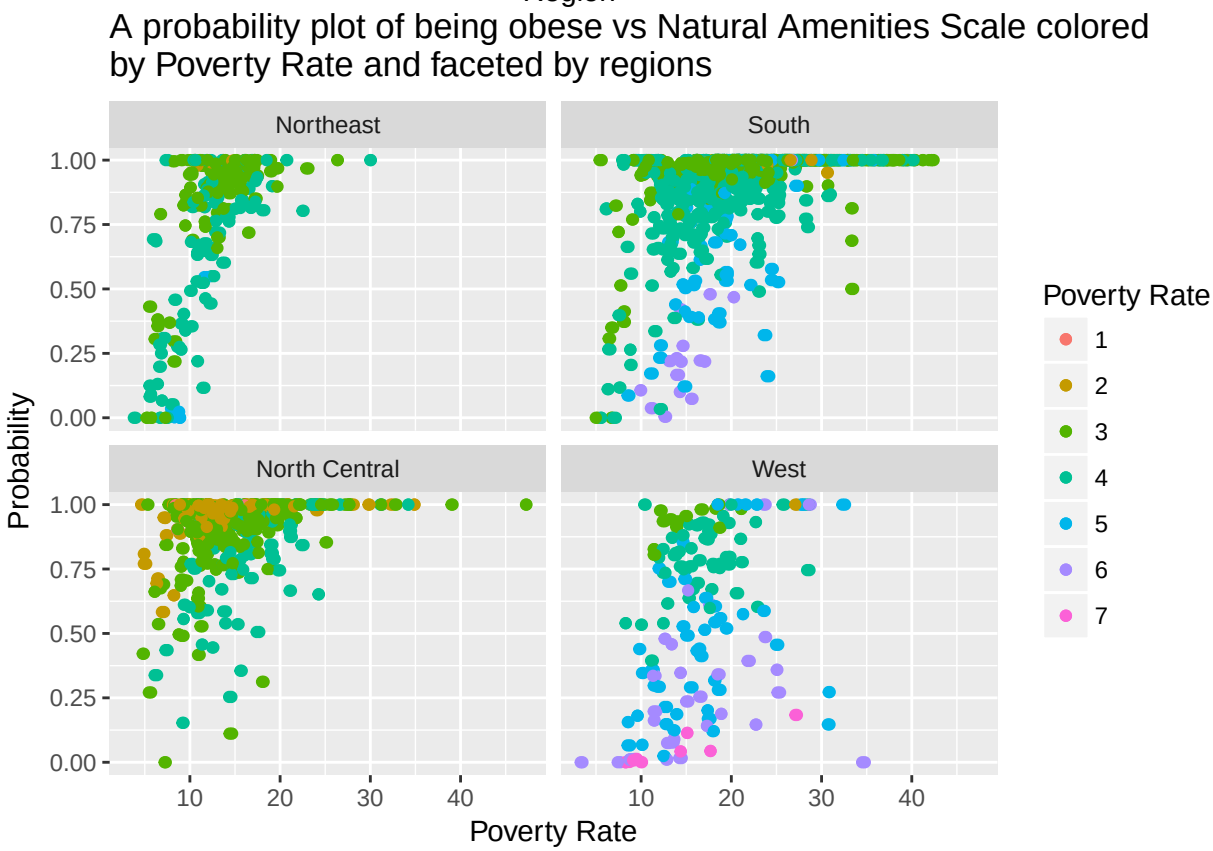


Plot of above vs Natural Amenities Scale

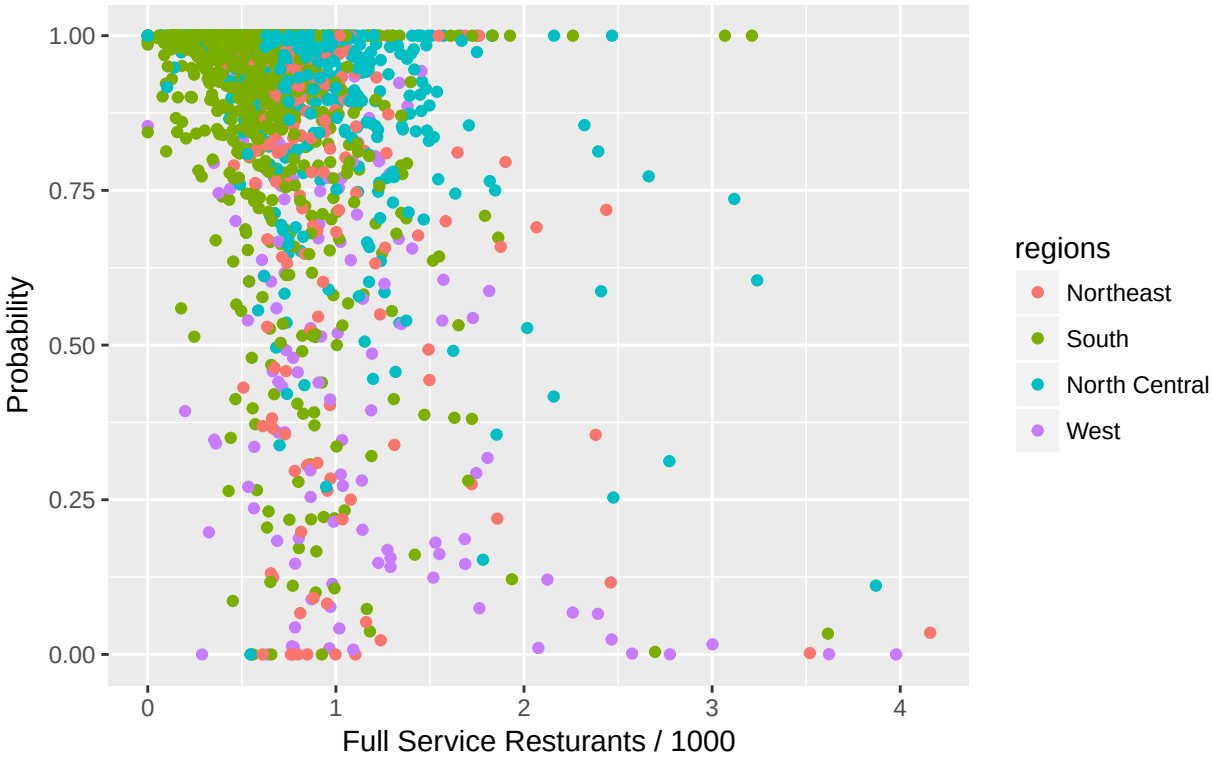



A probability plot of being obese vs Natural Amenities Scale colored by re
Probability
1.00
0.75
0.50
0.25
0.00
2
4
6
NATAMEN
Region
Northeast
South
North Central
West
Probability
1.00
0.75
0.50
0.25
0.00
Northeast
South
North Central
West
Region
NATAMEN
1
2
3
4
5
6
7

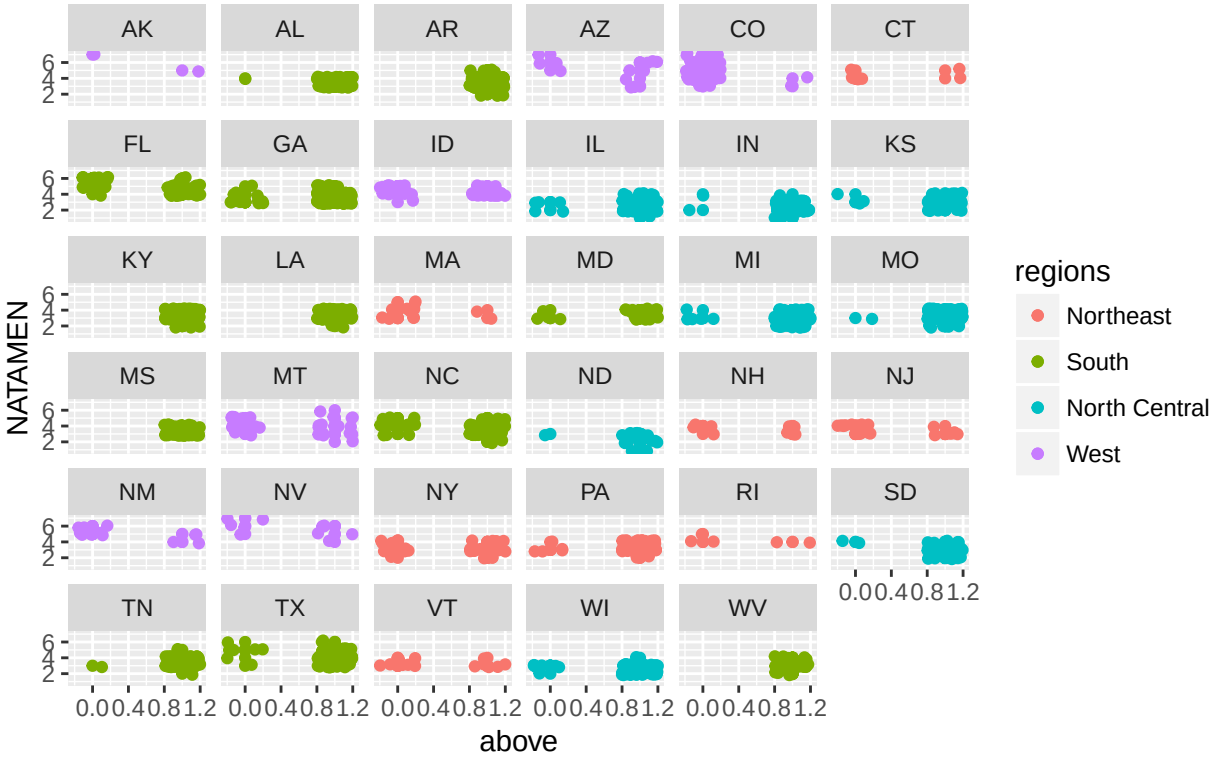

A probability plot of being obese vs regions colored by Natural Amenities S
Northeast
South
North Central
West
Probability
1.00
0.75
0.50
0.25
0.00
2
4
6
Region
NATAMEN
40
30
20
10


A probability plot of being obese vs Natural Amenities Scale colored by Poverty Rate and faceted by regions
Northeast
South
North Central
West
Probability
1.00
0.75
0.50
0.25
0.00
10
20
30
40
Poverty Rate
Poverty Rate
1
2
3
4
5
6
7

A probability plot of being obese vs
Full Service Resturants coloured by regions



A plot of Natural Amenities Scale vs above colored by
regions and faceted by states



Limitations

-We weren't sure which type of diabetes this measures. -Didn't report how they got the data. -This is an observational study so they had no control over the variables. We just looked for explanatory patterrs. -We had to take a a lot of NA and had to take out a bunch of data points (only 35 states left). -The income reported is the median income for each county and has an error margin of +- 5%. -We would have liked more categories for the information they had.

Conclusion

-We determined FSRPTH07, NATAMEN, and Poverty Rate were the three best variables to predict whether your county would be above and below obesity rate. -We made a good predictive model. -Don't live in Kentucky. -