

# Diabetes

## LEADING CONTRIBUTING FACTORS

Lesley Conn, Katrina Brown, Triston  
Cavaness and Juliet Hamilton



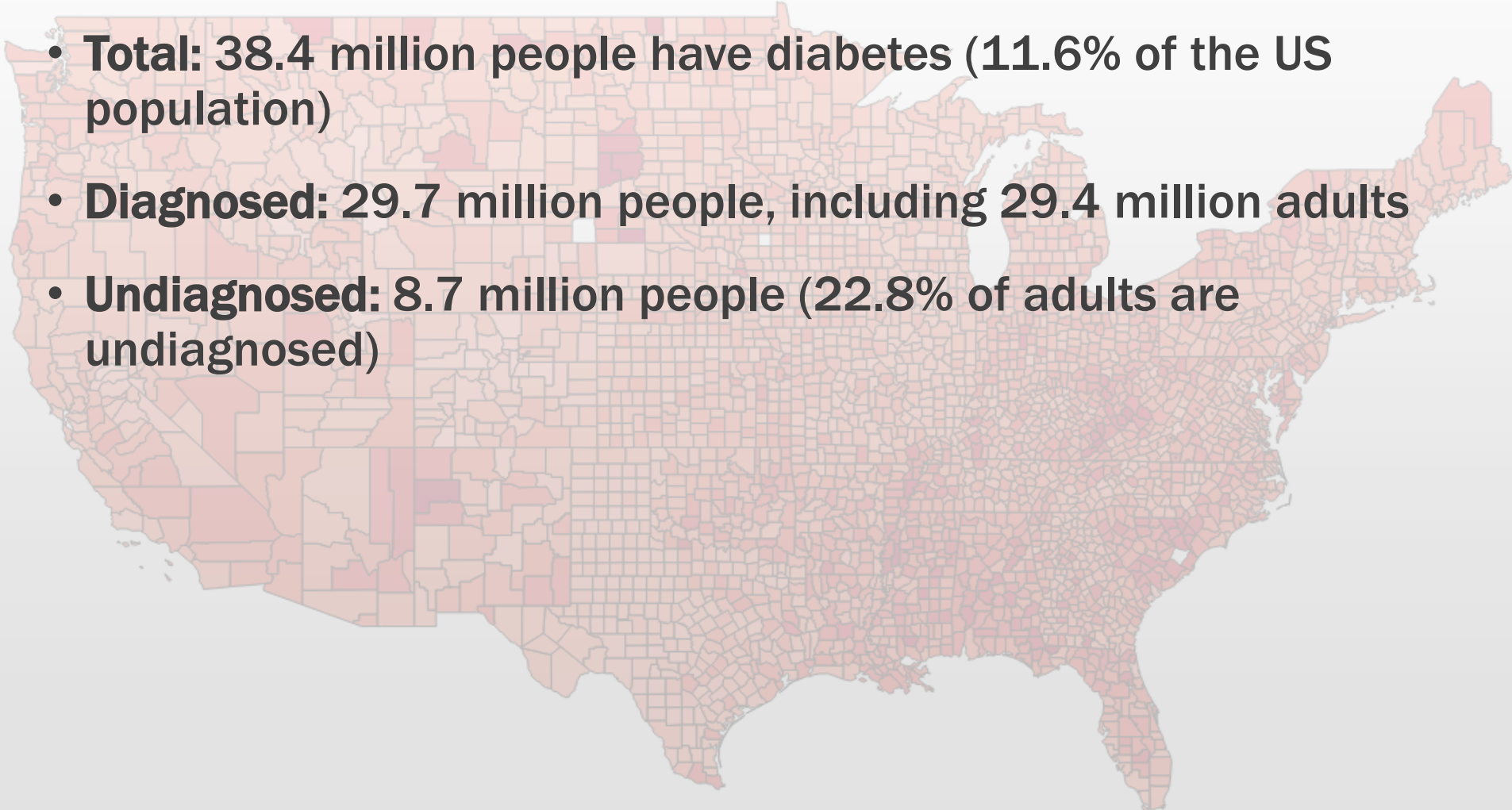
# Diabetes Onset and Diagnosis Review

- Data used
  - Kaggle data that contained 100,000 rows
  - CDC information provided 2018 research on diabetes and obesity for more than 3,141 counties in the United States
- Areas analyzed
  - Male and female populations
  - Age
  - Body Mass Index (BMI)
  - Smoking history
  - Hypertension
  - Fasting and non-fasting blood glucose levels
  - Heart disease
- Limitation
  - Neither source included patient race and/or ethnicity, though some hypotheses were developed after mapping the CDC data based on county codes, then referencing locations against U.S. Census data for those counties.

	gender	age	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes
0	Female	80.0	0	1	never	25.19	6.6	140	0
1	Female	54.0	0	0	No Info	27.32	6.6	80	0
2	Male	28.0	0	0	never	27.32	5.7	158	0
3	Female	36.0	0	0	current	23.45	5.0	155	0
4	Male	76.0	1	1	current	20.14	4.8	155	0
5	Female	20.0	0	0	never	27.32	6.6	85	0
6	Female	44.0	0	0	never	19.31	6.5	200	1
7	Female	79.0	0	0	No Info	23.86	5.7	85	0
8	Male	42.0	0	0	never	33.64	4.8	145	0
9	Female	32.0	0	0	never	27.32	5.0	100	0

# United States Diabetes Facts

- **Total:** 38.4 million people have diabetes (11.6% of the US population)
- **Diagnosed:** 29.7 million people, including 29.4 million adults
- **Undiagnosed:** 8.7 million people (22.8% of adults are undiagnosed)

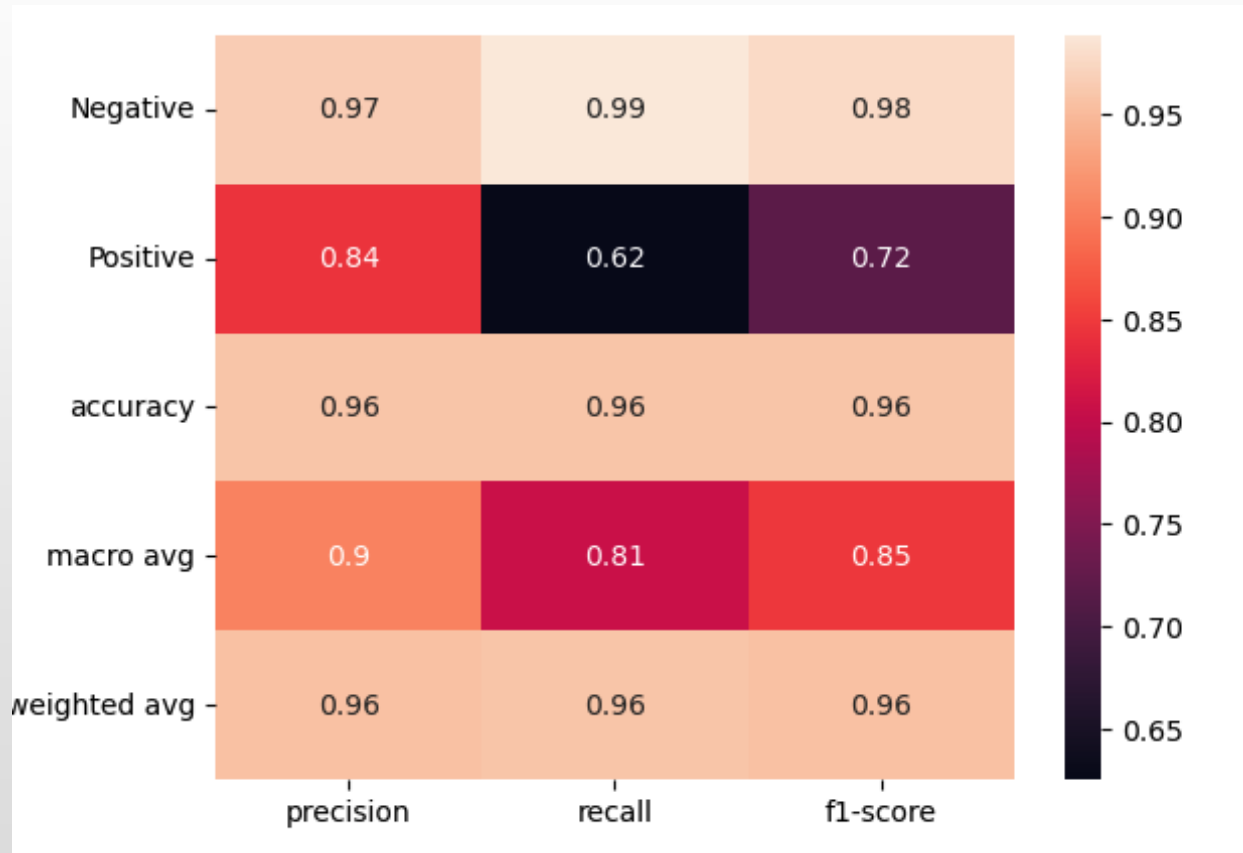


# Risk Factors for Diabetes-Related Complications

- Smoking
- Overweight and Obesity
- Physical Inactivity
- A1C
- High Blood Pressure
- High Cholesterol

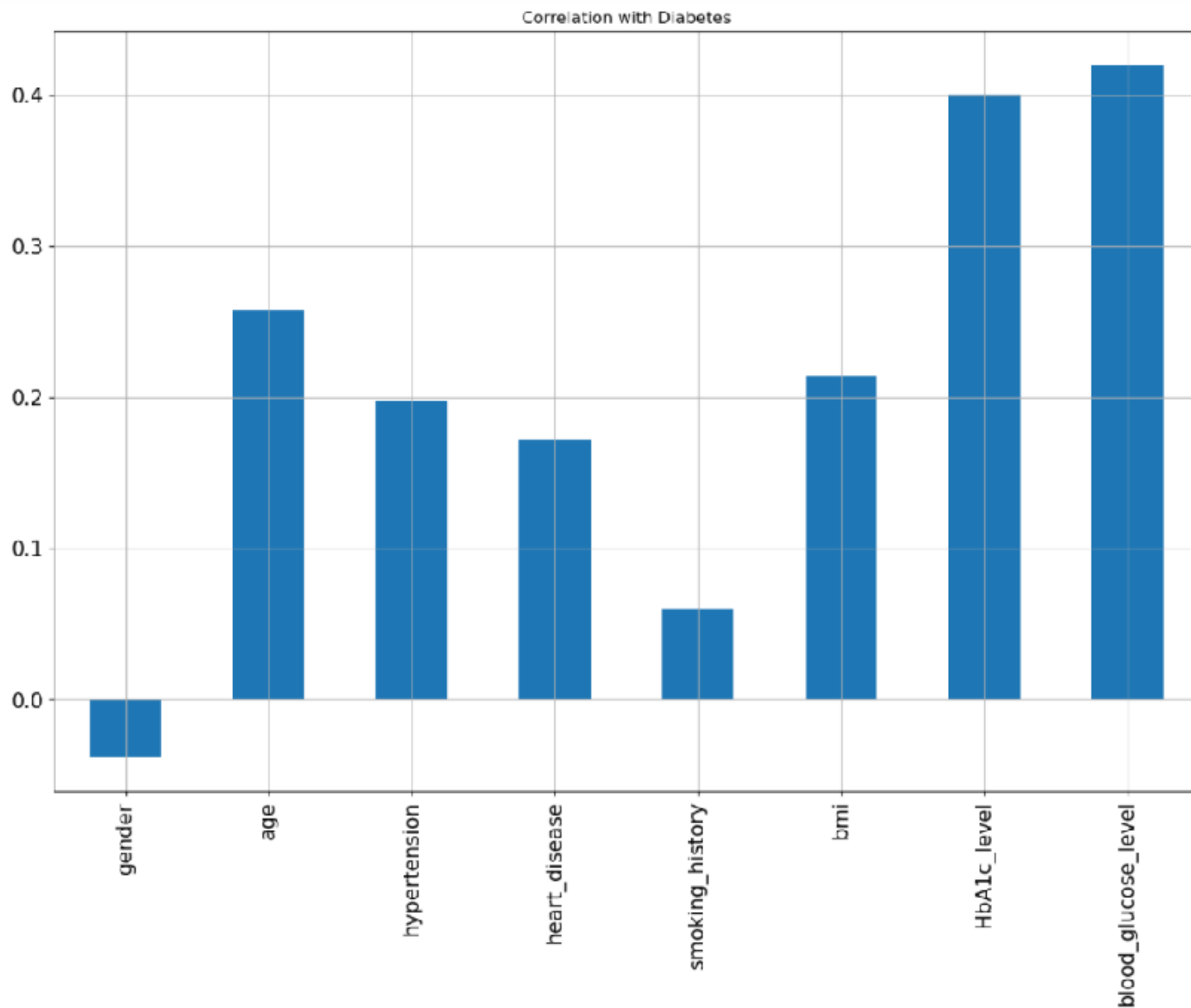






## Classification Report

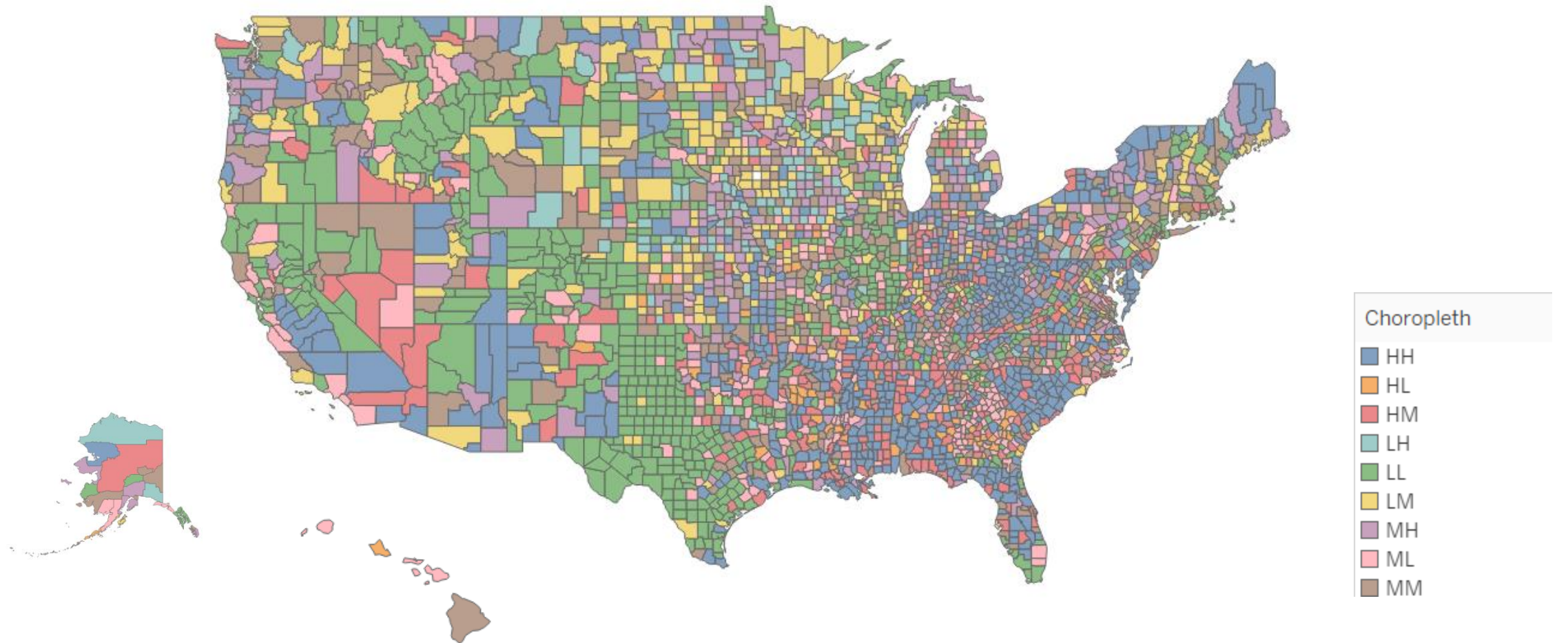
- The first model attained an accuracy score of 96% and F1 scores of 98% (Class 0 for detecting no diabetes) and 72% (Class 1 for detecting diabetes) on the harmonic of precision and recall.
- The noticeable difference in the precision and recall for Class 1 indicates that the model struggled more with correctly identifying instances of diabetes.



## Correlation with Diabetes

The Kaggle data examined gender, age, hypertension, heart disease, smoking history, BMI, HBA1C levels, and glucose levels.

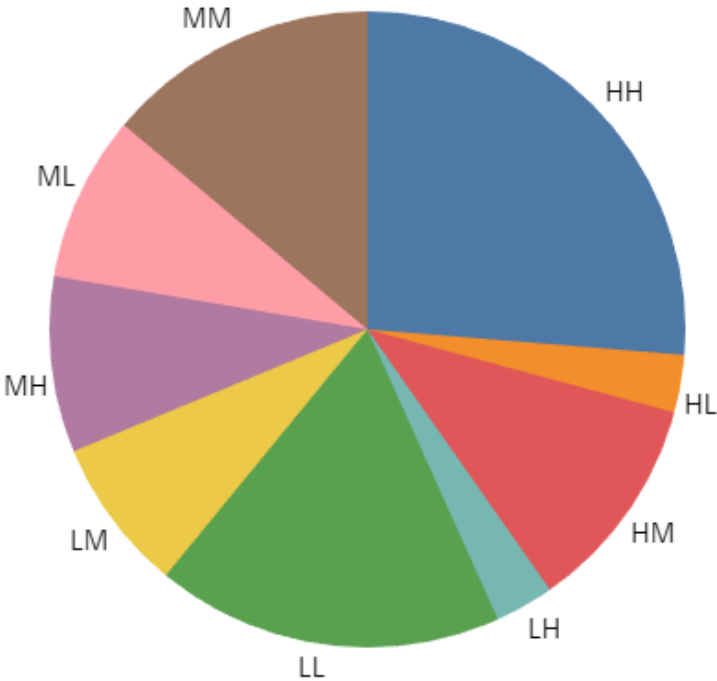
# Bivariate Choropleth for Diagnosed Diabetes and Obesity (2018)



Choropleth	
HH	
HL	
HM	
LH	
LL	
LM	
MH	
ML	
MM	

Summary	
Count:	9
SUM(Diagnosed Diabetes (%))	
Sum:	27,379
Average:	3,042
Minimum:	802
Maximum:	7,195
Median:	2,447
Standard deviation:	2,021
SUM(Obesity (%))	
Sum:	86,467
Average:	9,607
Minimum:	1,822
Maximum:	22,148
Median:	8,347
Standard deviation:	6,092



# Choropleth for Diagnosed Diabetes

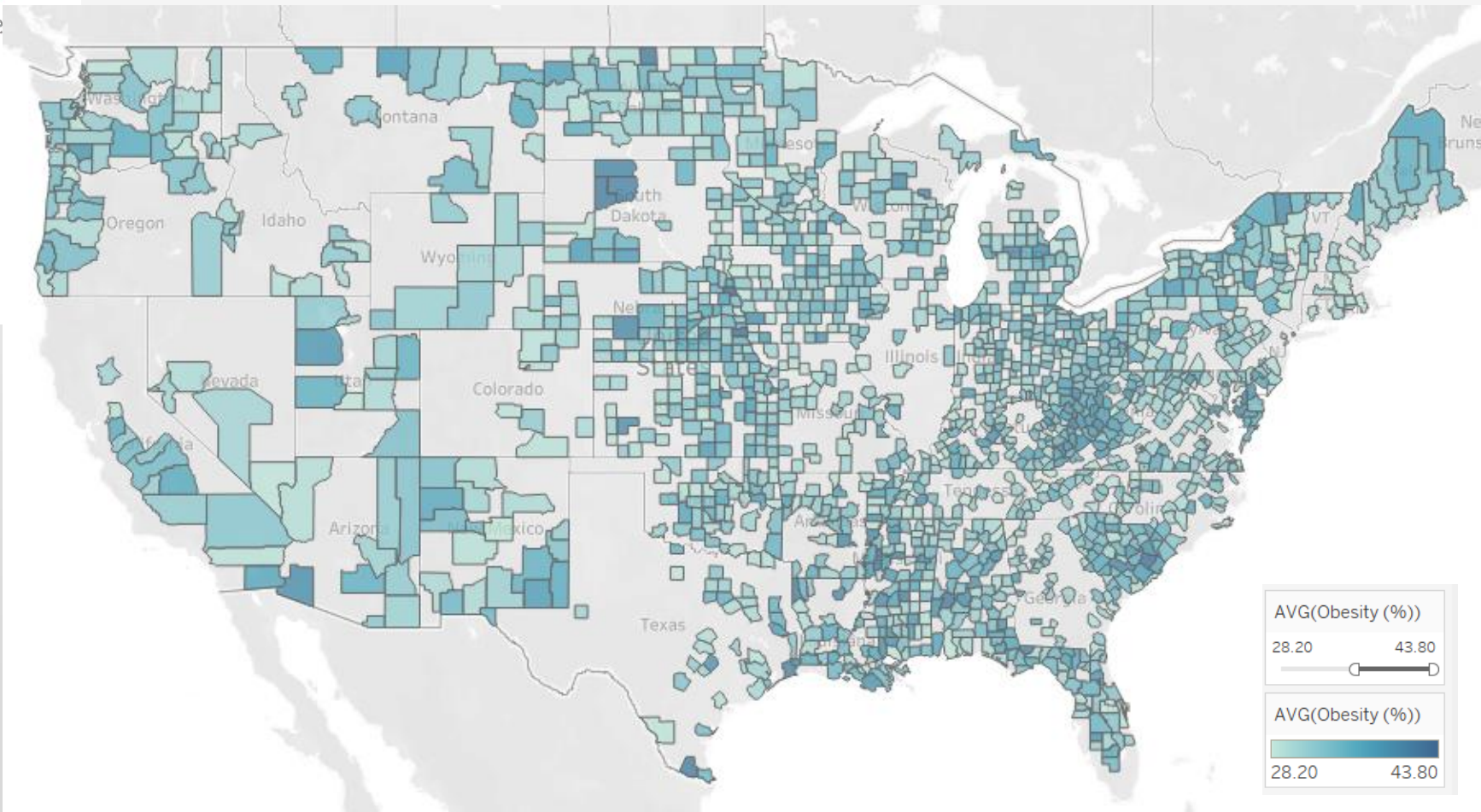
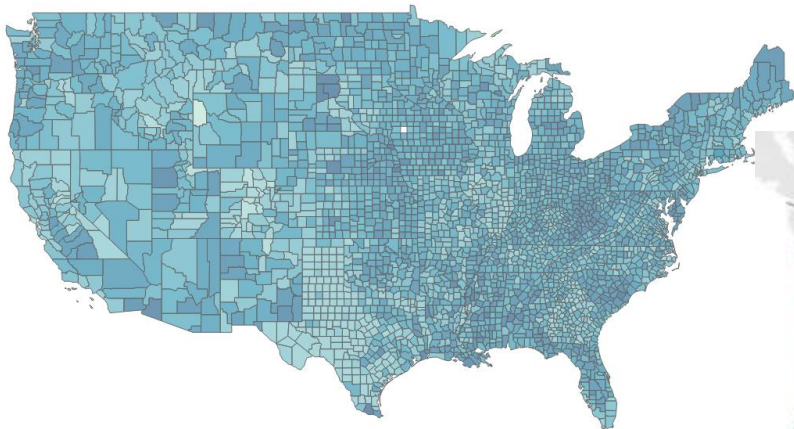
This analysis is based on Obesity and Diabetic results

HH represents the largest percentage with 7.2% Diabetic and 22.2% Obese

LL represents the second largest grouping with 4.8% Diabetic and 14.2% Obese

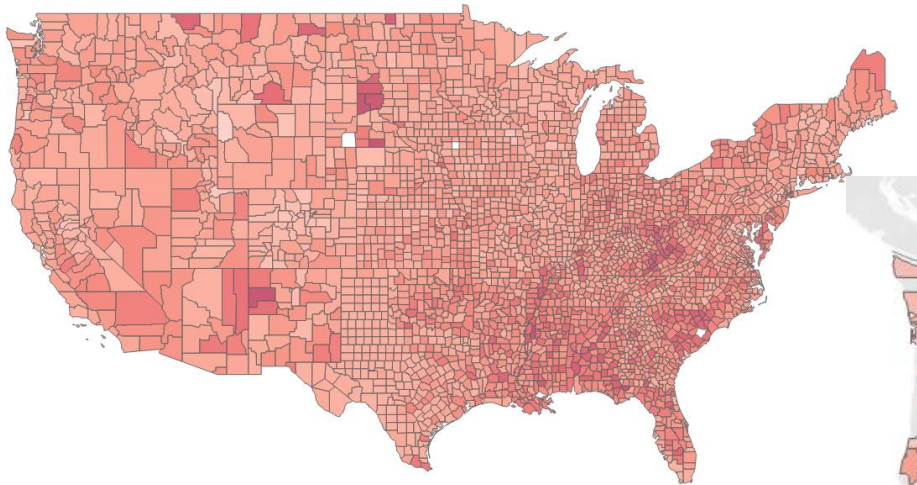


# Obesity Map by County Results

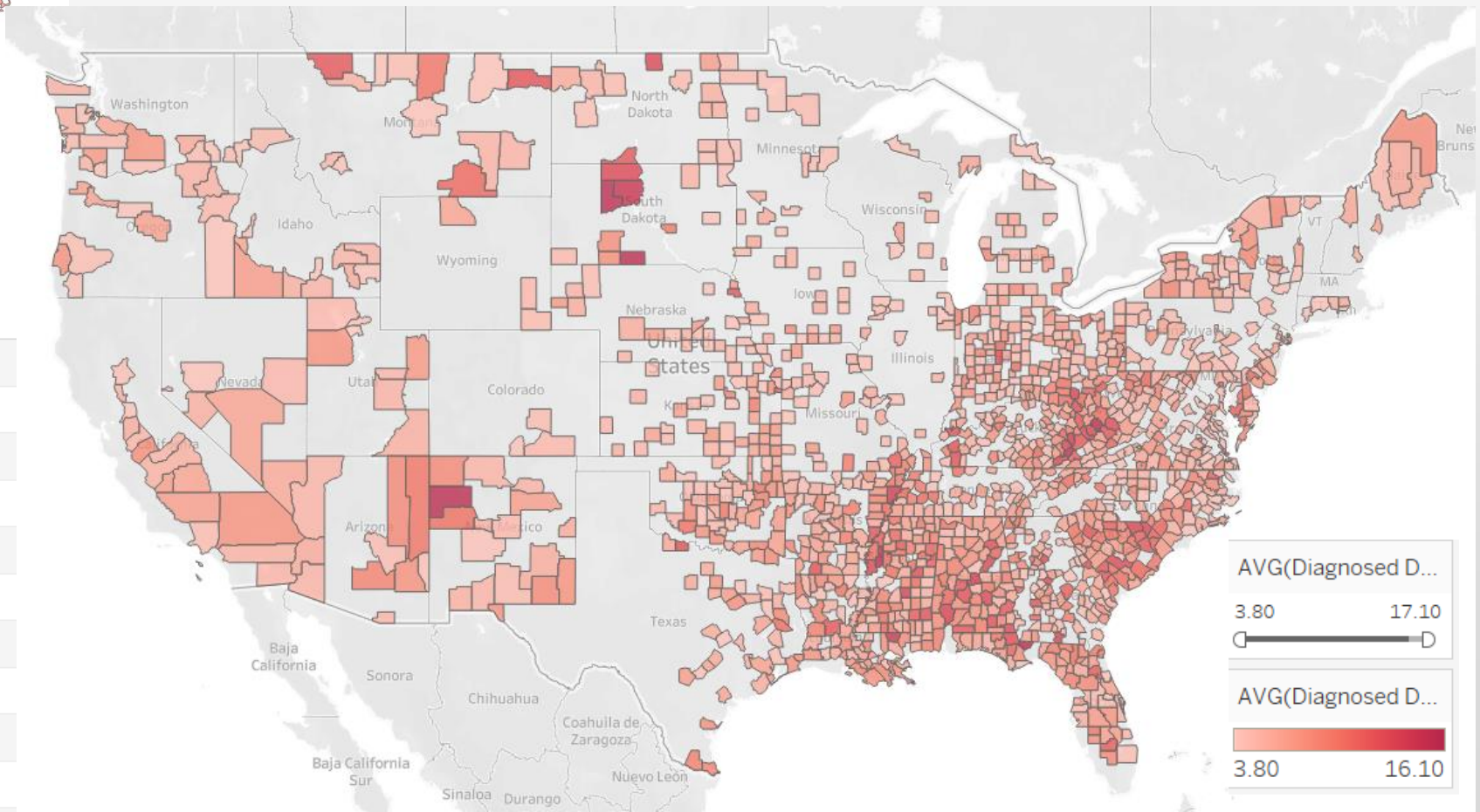


County	State	Avg. Obesity (%)
Ziebach County	South Dakota	41.5000
Yuma County	Arizona	39.6000
Young County	Texas	28.3000
York County	South Carolina	31.3000
York County	Nebraska	39.7000
York County	Maine	29.1000
Yellow Medicine County	Minnesota	30.8000
Yates County	New York	30.7000
Yankton County	South Dakota	31.5000
Yamhill County	Oregon	32.6000

# Diabetic Map by County Results



County	State	Avg. Diagnosed Diabete...
Ziebach County	South Dakota	15.9000
Zavala County	Texas	6.7000
Zapata County	Texas	6.8000
Yuma County	Colorado	6.8000
Yuma County	Arizona	9.5000
Yukon-Koyukuk Census Area	Alaska	9.9000
Yuba County	California	8.0000
Young County	Texas	8.3000
York County	Virginia	9.3000
York County	South Carolina	9.7000





# Preventing Diabetes-Related Complications

- Usual Source for Diabetes Care
- Physical Activity
- Weight Management
- Statin Treatment
- A1C, Blood Pressure, Cholesterol, and Smoking (ABCs)



# Questions

