A Population Health Informatics Approach for Measuring Obesity Prevalence among VHA Patients

Hadi Kharrazi *MD PhD*

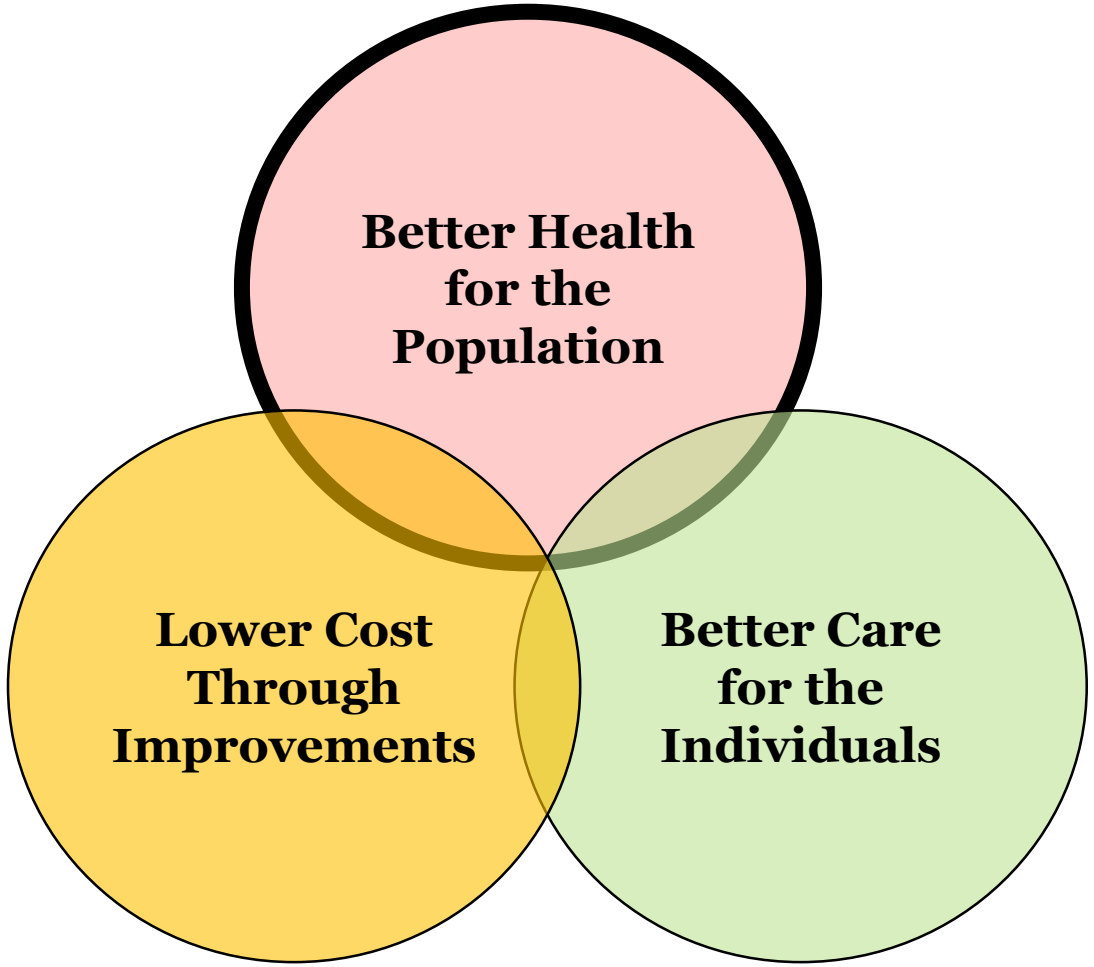
Johns Hopkins School of Public Health
Department of Health Policy and Management
Center for Population Health IT (CPHIT)
kharrazi@jhu.edu

Poll Question #1

- What is your primary role in VA? (select all that apply)
  - Student, trainee, or fellow
  - Clinician
  - Researcher, analyst, or data/IT scientist
  - Administrator, manager, or policy maker
  - Other

Population Health Informatics / IT

Population Health – An Emerging Field



Triple Aims developed by the Institute for Healthcare Improvement (IHI)

Population Health Informatics

JHU Center for Population Health Information Technology

(CPHIT)

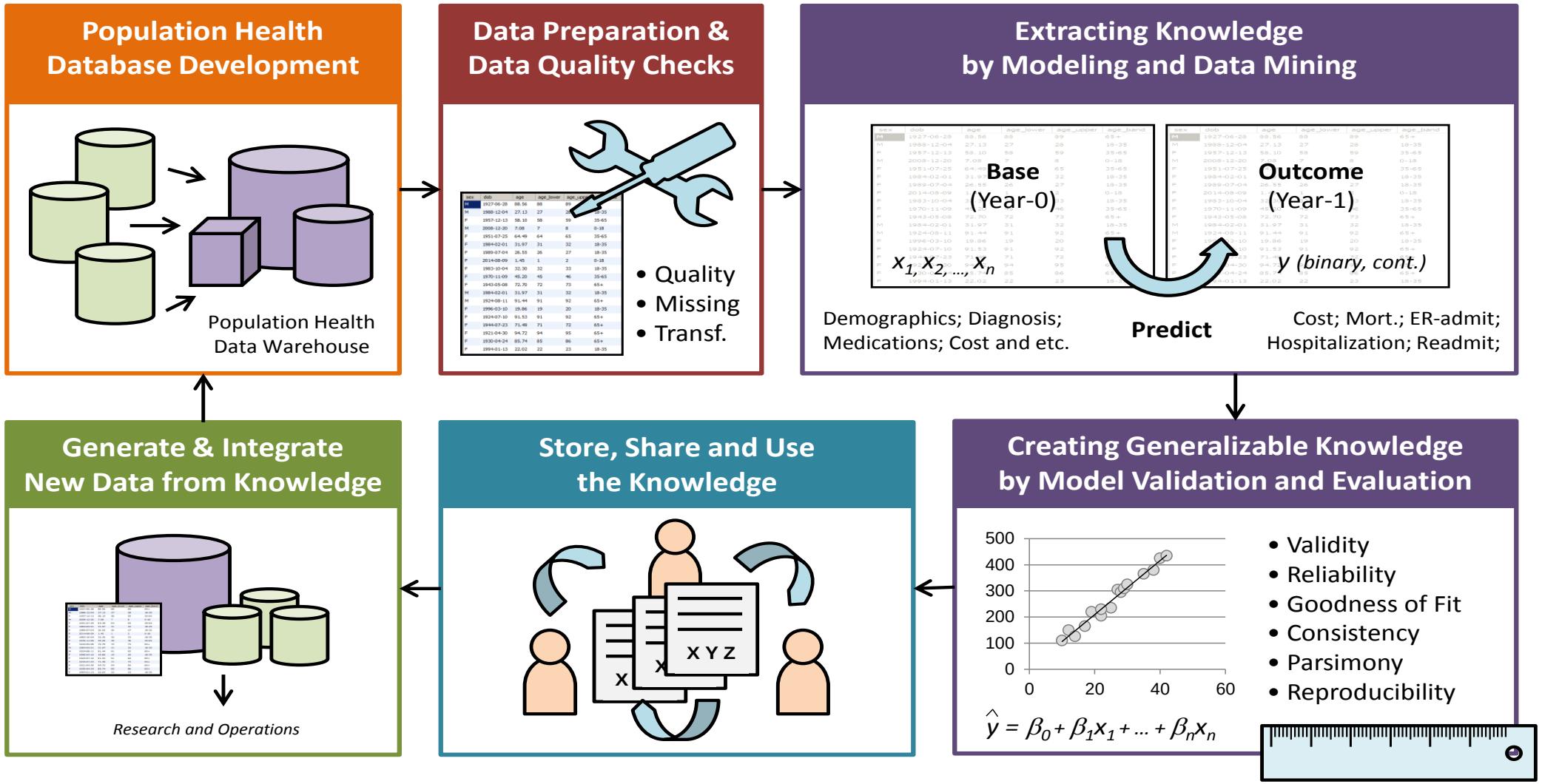
- *CPHIT improves the health of populations by advancing cutting edge health IT across all sectors*

- Outcomes: Healthcare Utilization (e.g., cost, hospitalization, ER admission)
- Predictors: Demographics, Diagnoses, Medications, Social Determinants + "new variables"
- Data Source: Insurance Claims, EHRs, HIEs, Hospital Discharges
- Scale: Populations (n = mil+)
- Temporal: Multi-year (t = 3 yrs+)

- Director: Dr. Weiner
- Research Director: Dr. Kharrazi

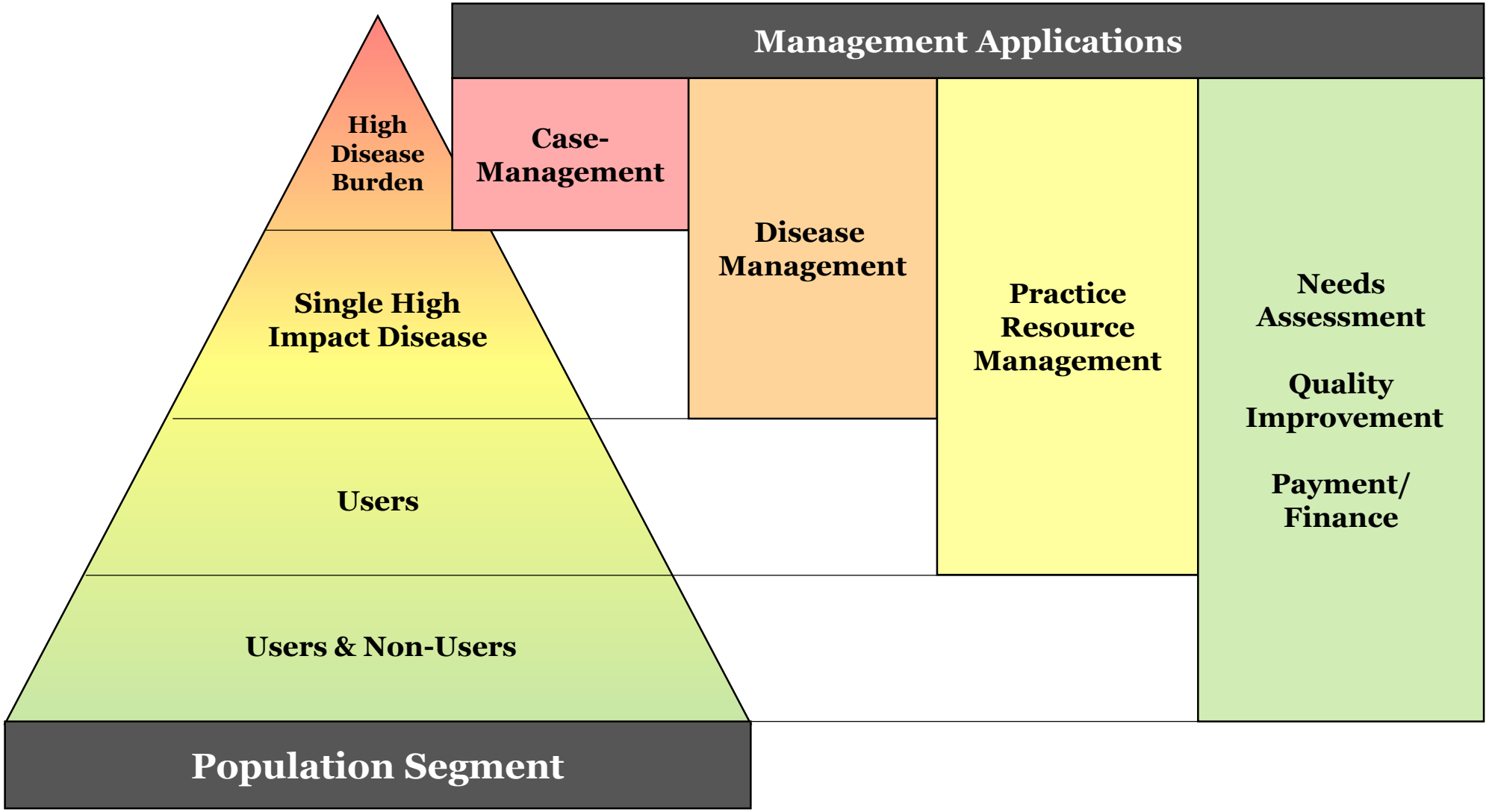
www.jhsph.edu/cphit

Data Analytic Cycle

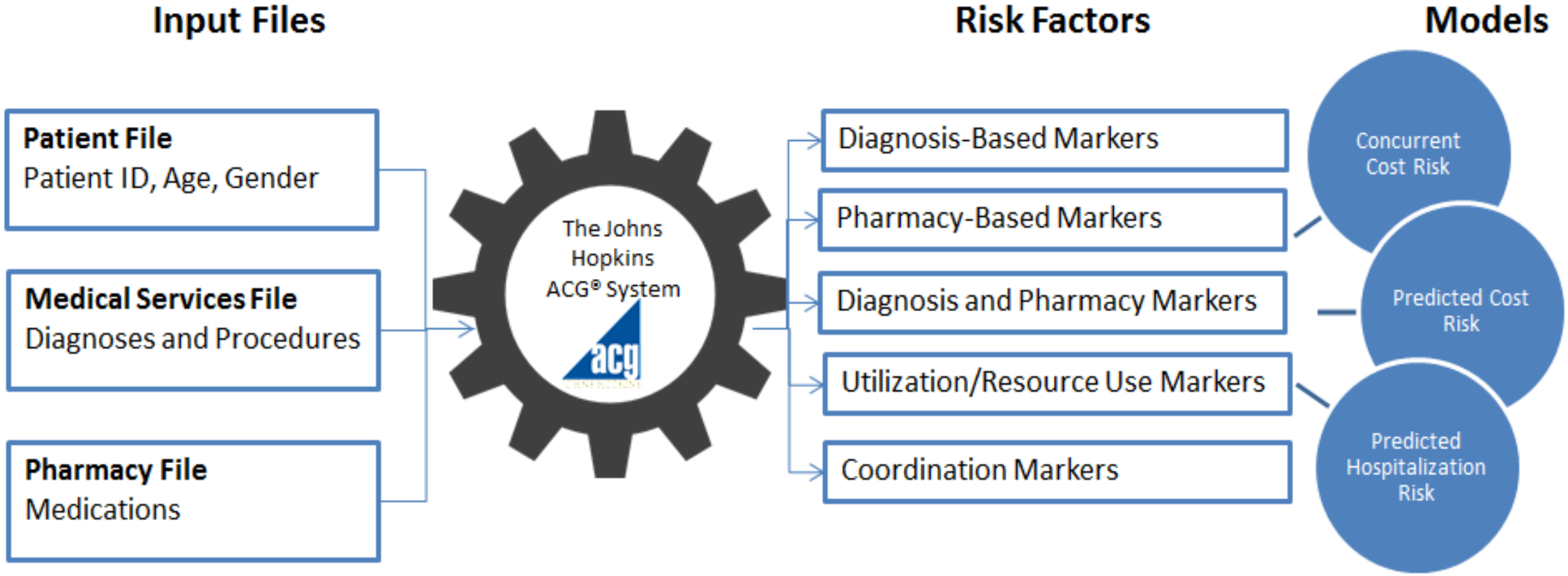


Overall Population Health Knowledge Management Process

Population Risk Stratification

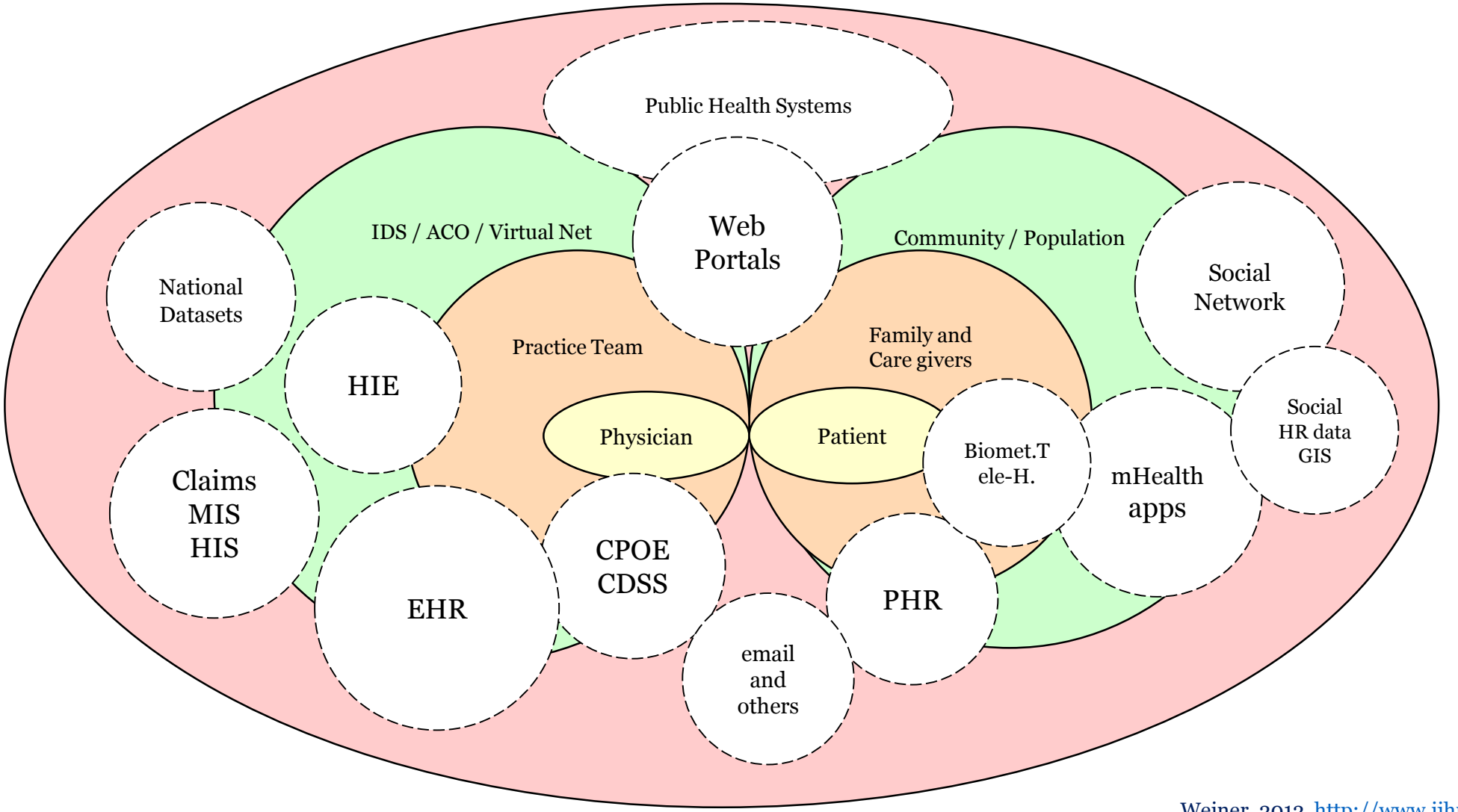


Population Risk Stratification



acg JOHNS HOPKINS

Population Data Sources



Weiner, 2012 http://www.ijhpr.org/content/1/1/33

CPHIT VHA Collaboration

- Developing a conceptual **population health analytic framework**
- Testing the **technical feasibility** of population health analytic framework within a population-level context (e.g., obesity as a population-level issue) using underlying existing data platforms (e.g., CDW, geo-data)
- Deriving **models to predict utilization** within the sample population denominator (i.e., trajectory among overweight/obese populations)
  - Exploring the prevalence of obesity on a temporal and geographic basis
- Expanding the geo-derived population-level analysis framework in evaluating utilization among VHA patients using **PACT services**

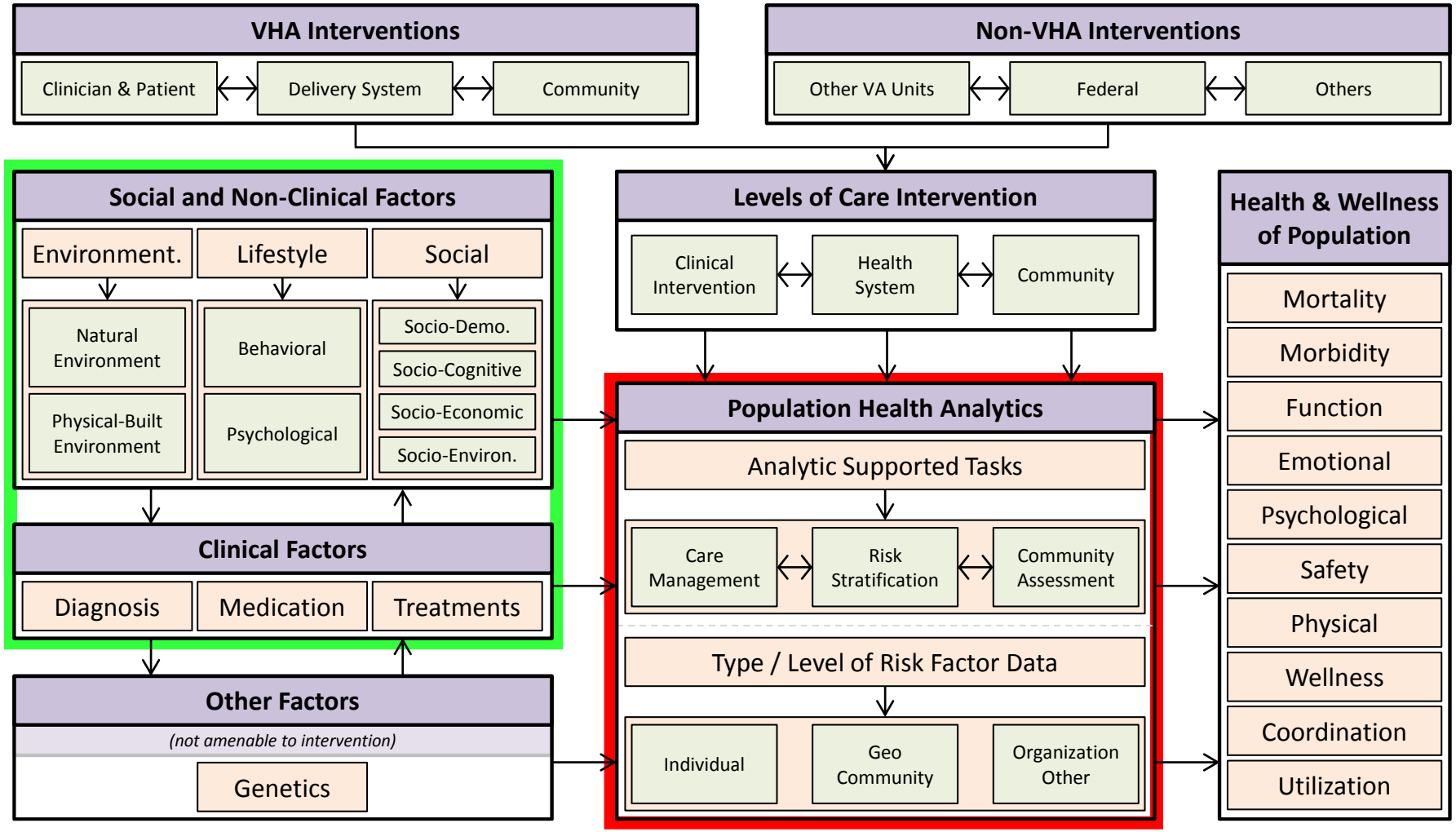
VHA Population Health Framework

" The Obesity Prevalence Project "

Poll Question #2

- What do you think about VHA patients' obesity rate compared to the general population? (select all that apply)
  - VHA population has a higher rate of obesity
  - VHA population has a different trend of obesity over time
  - VHA population has a different geo-distribution of obesity
  - VHA population has a different set of factors associated with obesity

VHA Population Health Conceptual Framework



Proposed VHA Population Health Framework

VHA Obesity Prevalence 2005

Obesity Prevalence Among Veterans at Veterans Affairs Medical Facilities

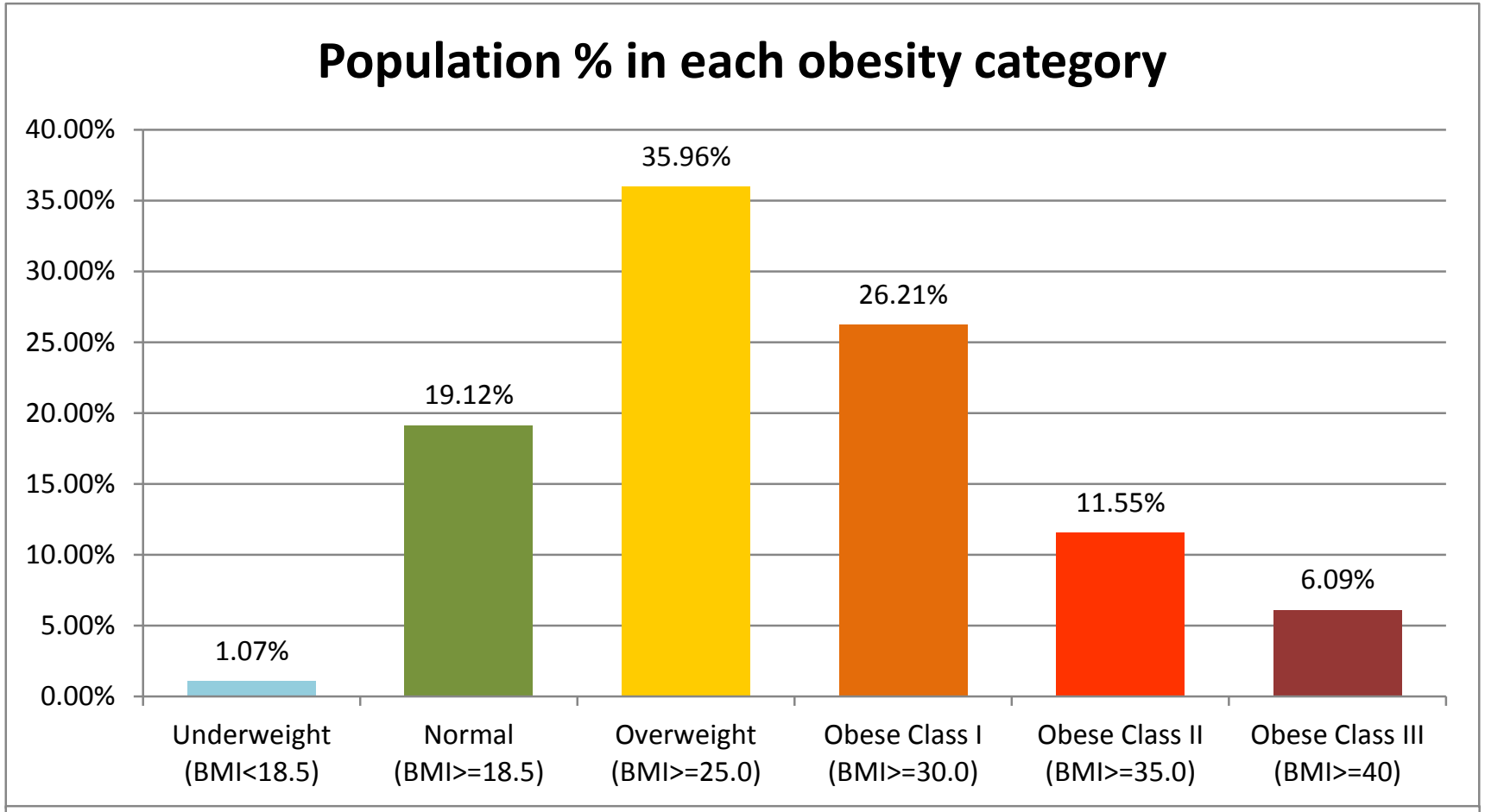
Sandeep R. Das, MD, MPH, Linda S. Kinsinger, MD, MPH, William S. Yancy Jr, MD, MHS, Anthea Wang, MD, MPH, Eileen Ciesco, MHA, Mary Burdick, PhD, RN, Steven J. Yevich, MD, MPH

Table 1. BMI category by age and stratified by gender[a]

	Overweight BMI 25–29.9 kg/m²		Class-I obese BMI 30–34.9 kg/m²		Class-II obese BMI 35–39.9 kg/m²		Class-III obese BMI≥40 kg/m²	
Gender and age (years)	***n***	**% (SE)**	***n***	**% (SE)**	***n***	**% (SE)**	***n***	**% (SE)**
Women (*n*=93,290)								
≥18	28,916	31.0 (0.01)	19,750	21.2 (0.01)	9,507	10.2 (0.01)	5,608	6.0 (0.02)
18–29	2,330	30.0 (0.02)	1,168	15.0 (0.03)	436	5.6 (0.05)	150	1.9 (0.08)
30–39	5,211	31.5 (0.02)	3,458	20.9 (0.02)	1,547	9.4 (0.03)	702	4.2 (0.04)
40–49	8,360	30.3 (0.01)	6,288	22.8 (0.01)	3,222	11.7 (0.02)	1,940	7.0 (0.02)
50–59	5,172	30.2 (0.02)	4,068	23.7 (0.02)	2,208	12.9 (0.02)	1,595	9.3 (0.03)
60–69	3,046	30.8 (0.02)	2,328	23.5 (0.02)	1,199	12.1 (0.03)	814	8.2 (0.04)
70–79	3,377	34.4 (0.02)	1,834	18.7 (0.03)	699	7.1 (0.04)	340	3.5 (0.06)
≥80	1,420	31.3 (0.03)	606	13.3 (0.04)	196	4.3 (0.07)	67	1.5 (0.12)
Men (*n*=1,710,032)								
≥18	685,909	40.1 (0.00)	379,112	22.2 (0.00)	126,918	7.4 (0.00)	55,905	3.3 (0.01)
18–29	7,861	39.0 (0.01)	3,885	19.3 (0.02)	1,192	5.9 (0.03)	368	1.8 (0.05)
30–39	24,264	38.5 (0.01)	15,324	24.3 (0.01)	5,652	9.0 (0.01)	2,627	4.2 (0.02)
40–49	70,853	36.5 (0.00)	45,543	23.5 (0.01)	18,095	9.3 (0.01)	9,666	5.0 (0.01)
50–59	142,534	36.8 (0.00)	95,677	24.7 (0.00)	38,349	9.9 (0.01)	20,565	5.3 (0.01)
60–69	157,412	40.3 (0.00)	96,937	24.8 (0.00)	33,367	8.5 (0.01)	14,027	3.6 (0.01)
70–79	215,680	43.7 (0.00)	100,709	20.4 (0.00)	26,247	5.3 (0.01)	7,822	1.6 (0.01)
≥80	67,305	41.7 (0.01)	21,037	13.0 (0.01)	4,016	2.5 (0.02)	830	0.5 (0.03)

[a]Data for underweight and normal weight individuals are not shown; however, indicated percentages are calculated including these individuals.
BMI, body mass index; SE, standard error.

Quick Prevalence Check



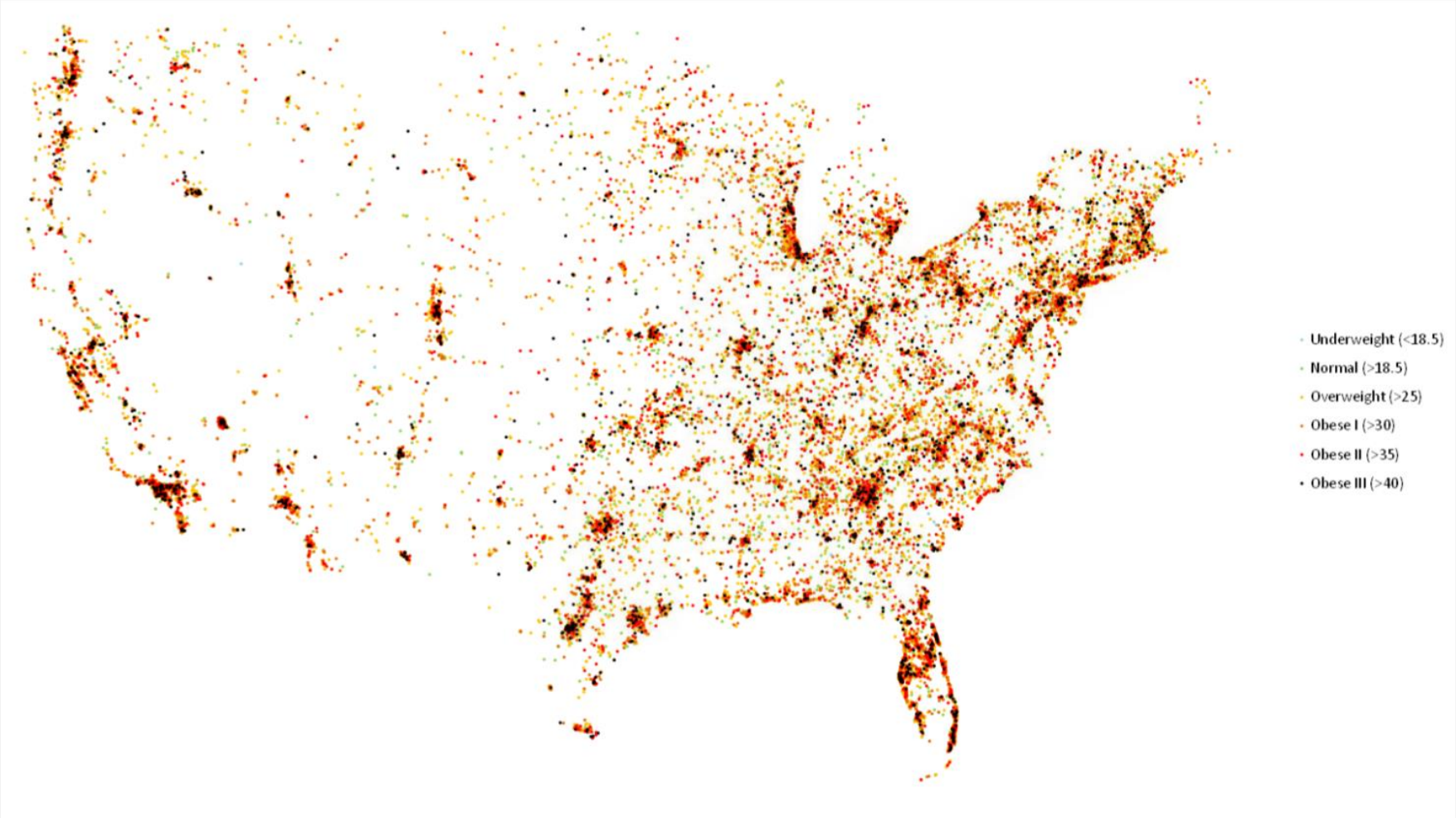
Average BMI per obesity category

(Limited to 29,322 visits occurred in one day of 2013; generated using CDW data)

50

46.82

Quick Data Coverage Check



Geographic distribution of obesity among VHA population

(Limited to 29,322 visits occurred in one day of 2013; generated using CDW data)

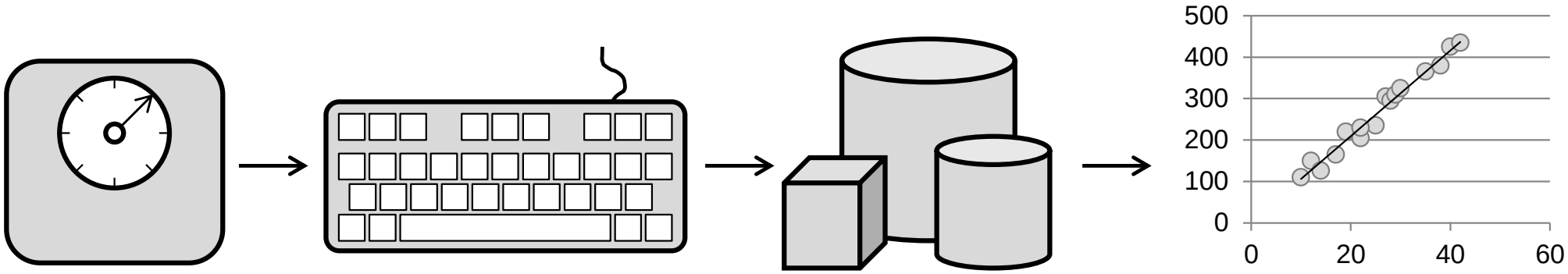
Data Quality Checks

Actual value:
200.6 lbs.

Recorded value:
200 lbs.

Data warehouse value:
200 kg

Analytic value:
100 kg (mean 200 & 0)



Measured (same day)
• Validity challenge
198.9 | 198.9 | 198.9 lbs.
• Reliability challenge
200.6 | 198.9 | 202.2 lbs.

Measured (diff. days)
• Missing data challenge
NULL | 200.6 | NULL lbs.

User Typed (one entry)
• Typos
200.6 lbs. → 20.06, 2006
• Mismatching units
200.6 lbs. → 200.6 kg
• Assumptions/truncations
200.6 lbs. → 200 lbs.
NULL → 0
• Free-text additions
200.6 lbs. → 200.6 pounds

DB Operations (one entry)
• Truncations/Rounding
200.6 → 200.0
• Error conversions
200.6 pounds → NULL
200.6 lbs. → 200.6 kg
• Cleaning
200+ lbs. → 200.0

Analytics (data points)
• Aggregation of data points
200 | 0 → mean of 100
• Selecting a representative
190 | 200 | 210 → 210 (first)
190 | 200 | 210 → 200 (mean)
190 | 200 | 210 → 210 (last)
• Removing outliers
200 | 200 | 350 → 200 | 200 | NULL

Example of data quality issues across various data collection and use stages

Cross-Sectional Obesity Prevalence 2015

	Normal weight BMI 18 24.9kg/m²			Overweight BMI 25 29.9 kg/m²			Class I obese BMI 20 24.9kg/m²			Class II obese BMI 35 39.9kg/m²			Class III obese BMI > 40 kg/m²		
Ages (years)	n	%	(SE)	n	%	(SE)	n	%	(SE)	n	%	(SE)	n	%	(SE)
Women															
18 29	13970	37.8	(0.25)	12029	32.6	(0.24)	6763	18.3	(0.2)	2832	7.7	(0.14)	997	2.7	(0.08)
30 39	21022	26.9	(0.16)	24058	30.8	(0.17)	18350	23.5	(0.15)	9586	12.3	(0.12)	4655	6.0	(0.08)
40 49	14232	19.2	(0.14)	22371	30.1	(0.17)	19819	26.7	(0.16)	11149	15.0	(0.13)	6475	8.7	(0.1)
50 59	20308	20	(0.13)	30108	29.6	(0.14)	26385	26.0	(0.14)	14812	14.6	(0.11)	9288	9.1	(0.09)
60 69	14390	20.6	(0.15)	20275	29.0	(0.17)	17824	25.5	(0.16)	9796	14.0	(0.13)	6901	9.9	(0.11)
70 79	3770	23.8	(0.34)	4915	31.0	(0.37)	3732	23.5	(0.34)	1908	12.0	(0.26)	1310	8.3	(0.22)
> 80	5079	42.8	(0.45)	3810	32.1	(0.43)	1700	14.3	(0.32)	591	5.0	(0.2)	223	1.9	(0.13)
Total	92771	23.9	(0.07)	117566	30.3	(0.07)	94573	24.3	(0.07)	50674	13	(0.05)	29849	7.7	(0.04)
Men															
18 29	39072	26.3	(0.11)	57568	38.8	(0.13)	33837	22.8	(0.11)	12945	8.7	0.07)	4579	3.1	(0.04)
30 39	50419	16.6	(0.07)	112106	36.8	(0.09)	86594	28.4	(0.08)	37897	12.4	0.06)	17096	5.6	(0.04)
40 49	39372	10.9	(0.05)	118323	32.8	(0.08)	115017	31.9	(0.08)	56265	15.6	0.06)	30923	8.6	(0.05)
50 59	102168	15.9	(0.05)	216302	33.6	(0.06)	185698	28.8	(0.06)	86971	13.5	0.04)	49835	7.7	(0.03)
60 69	257224	17.0	(0.03)	539426	35.7	(0.04)	421348	27.9	(0.04)	185723	12.3	0.03)	98517	6.5	(0.02)
70 79	169569	19.0	(0.04)	355021	39.7	(0.05)	236864	26.5	(0.05)	89336	10.0	0.03)	38158	4.3	(0.02)
> 80	237528	33.0	(0.06)	311925	43.3	(0.06)	125818	17.5	(0.04)	30450	4.2	0.02)	7606	1.1	(0.01)
Total	895352	19.5	(0.02)	1710671	37.3	(0.02)	1205176	26.3	(0.02)	499587	10.9	0.01)	246714	5.4	(0.01)

Methods

- Data Quality checks (completeness, accuracy, timeliness)
- Visualization and age adjustment**:
  - LMM Linear Model Regression + Mean Age
  - LMR Linear Model Regression + Random Age
  - MLM* Multi-level Model
  - EPI-DIR* Epidemiological Direct Adjustment
  - EPI-IND* Epidemiological Indirect Adjustment

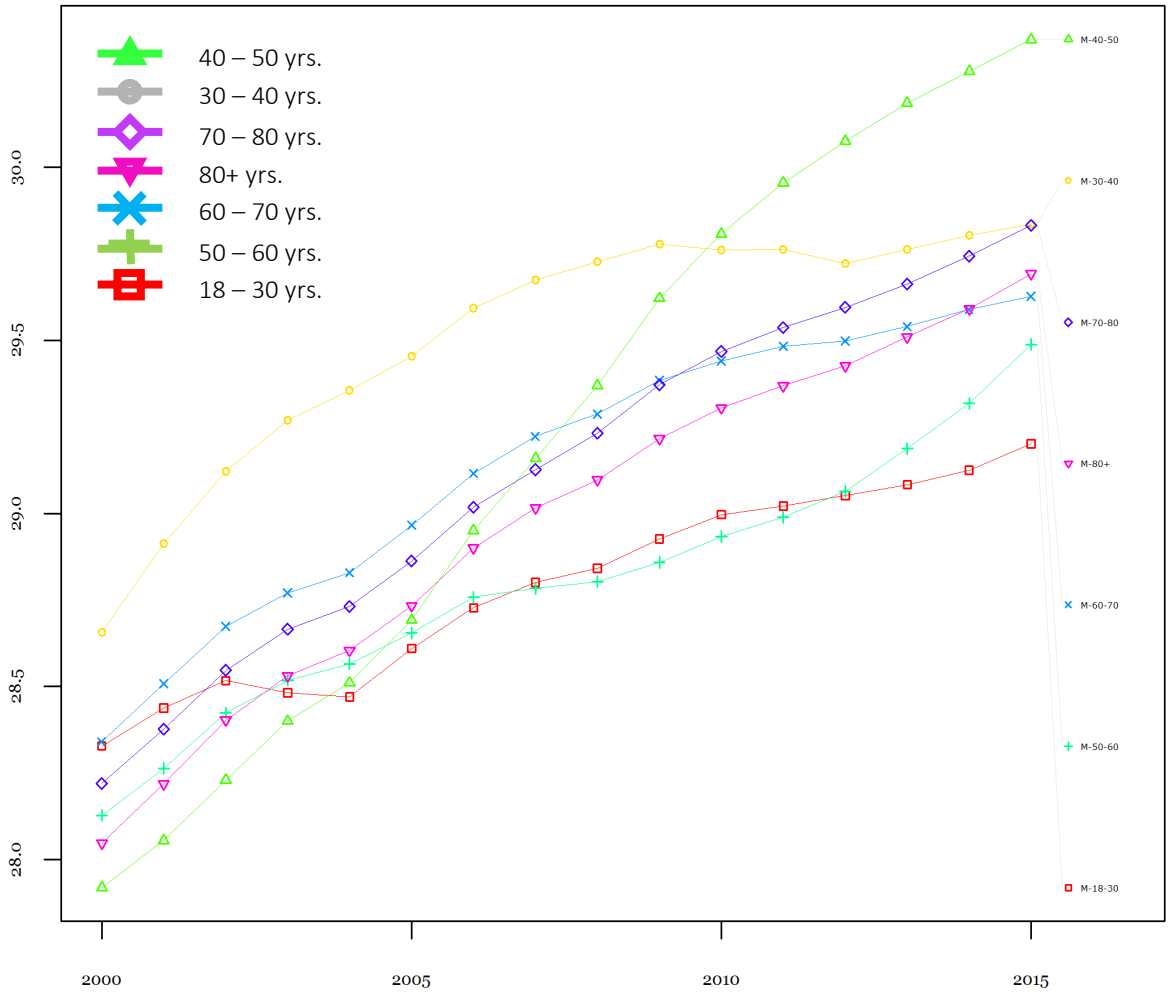
** control for geographical differences*

*** levels: State, County, Tract, Zip, and Block Group*

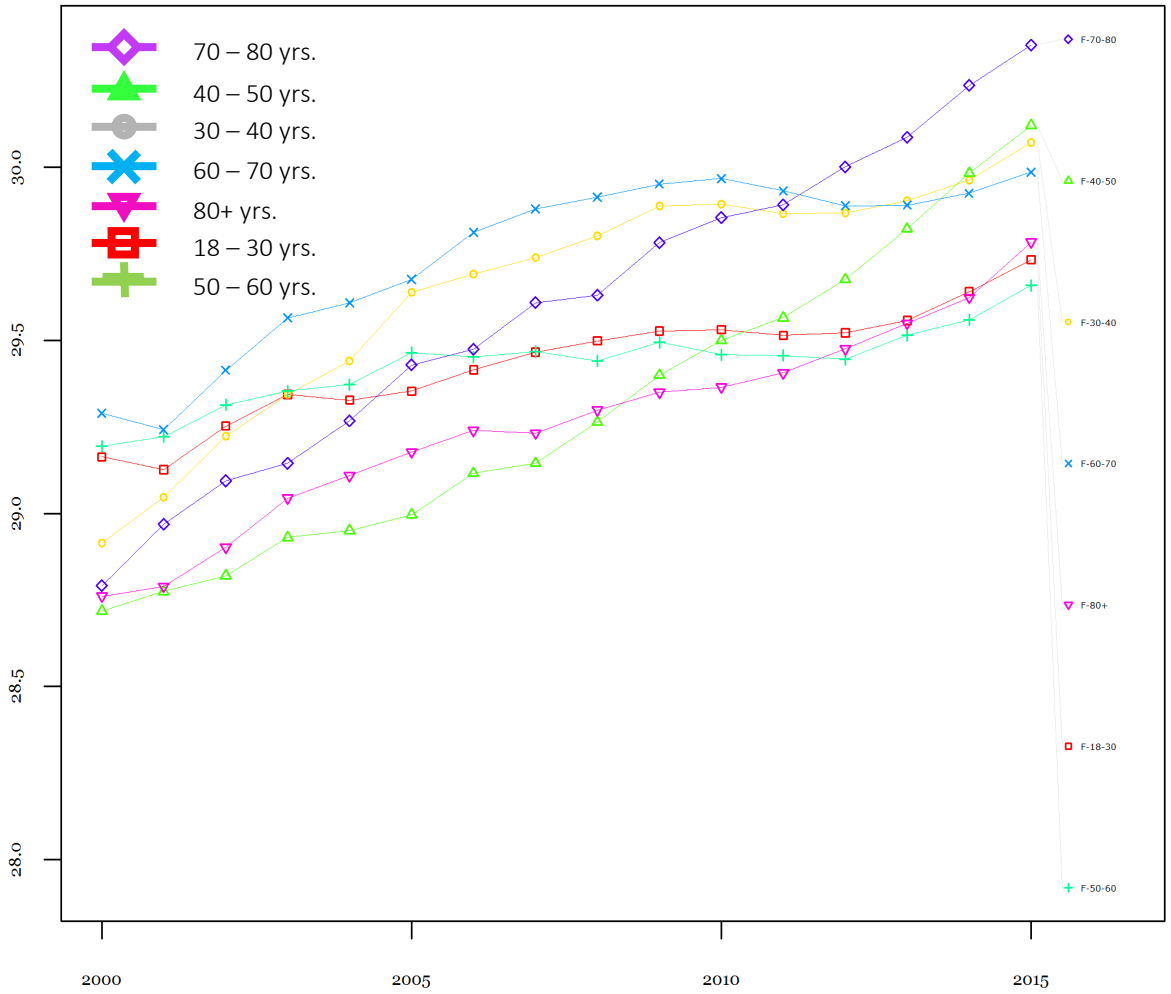
- GIS cluster analysis:
  - Poisson modeling
  - Bernoulli analysis
- Low sample size for female veterans → possibility of skewed/biased results (consider obesity rates among female veterans as preliminary)
- <u>Results are unpublished – changes might occur after peer-review</u>

Age & BMI [LMR] 2000-2015

Male

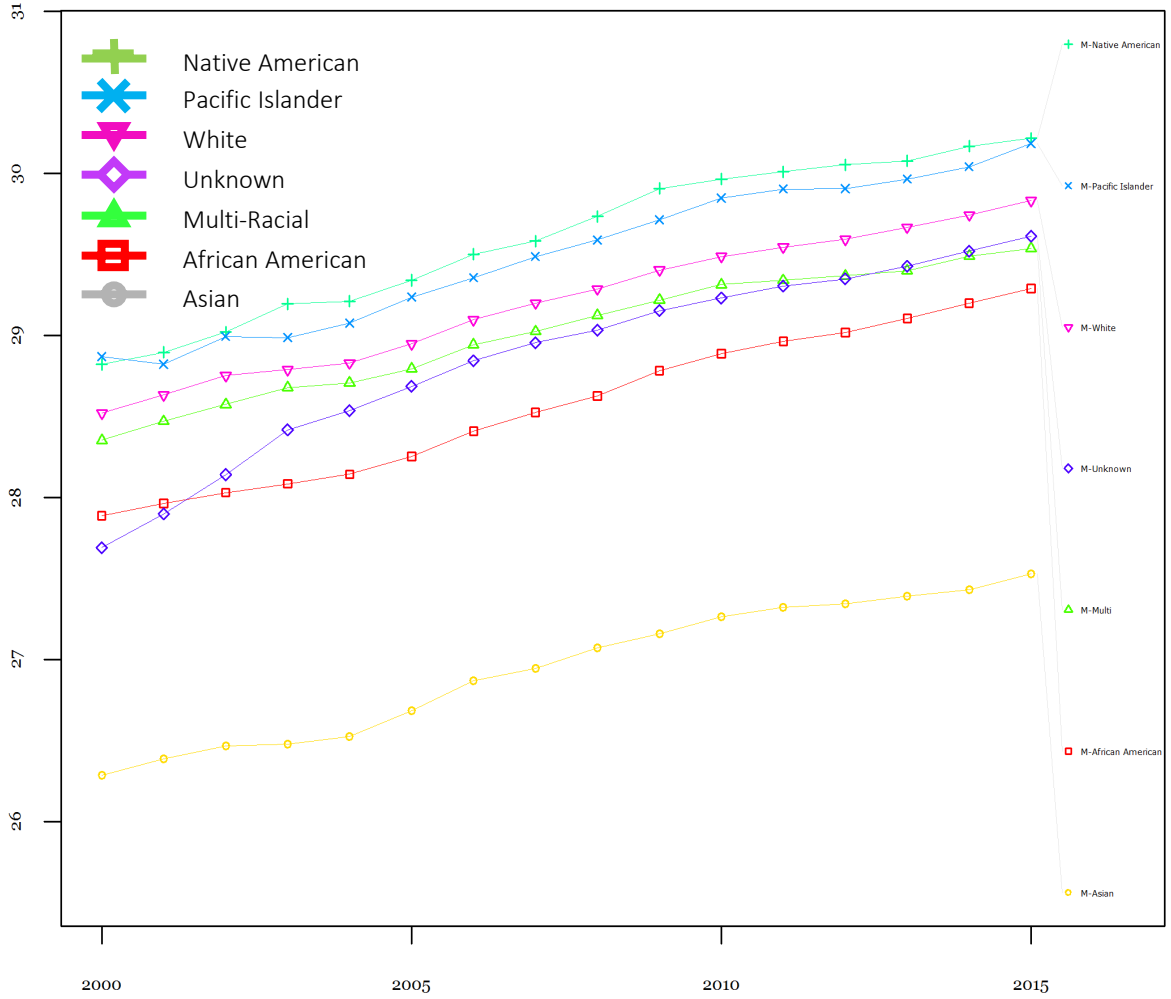


Female

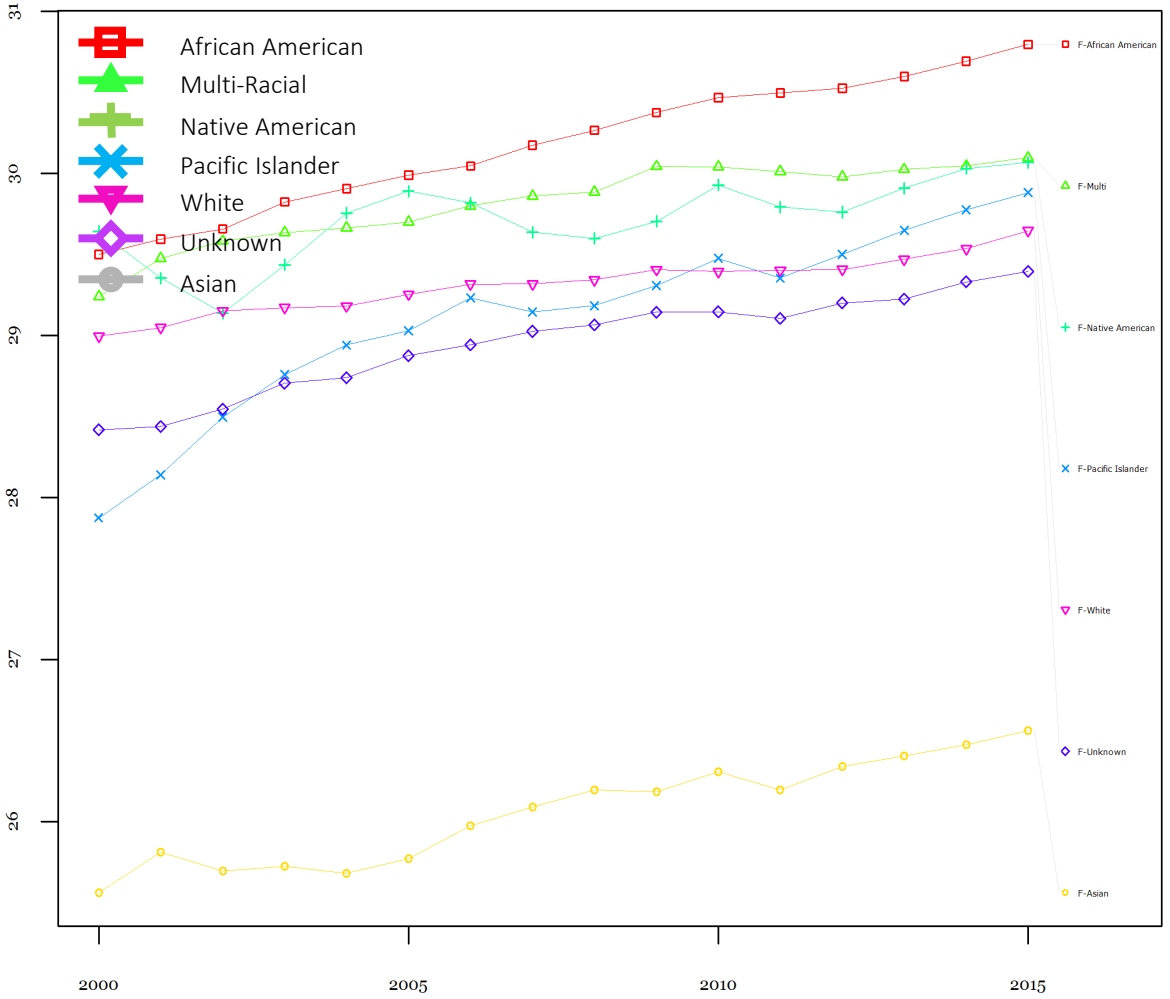


Race & BMI [LMR] 2000-2015

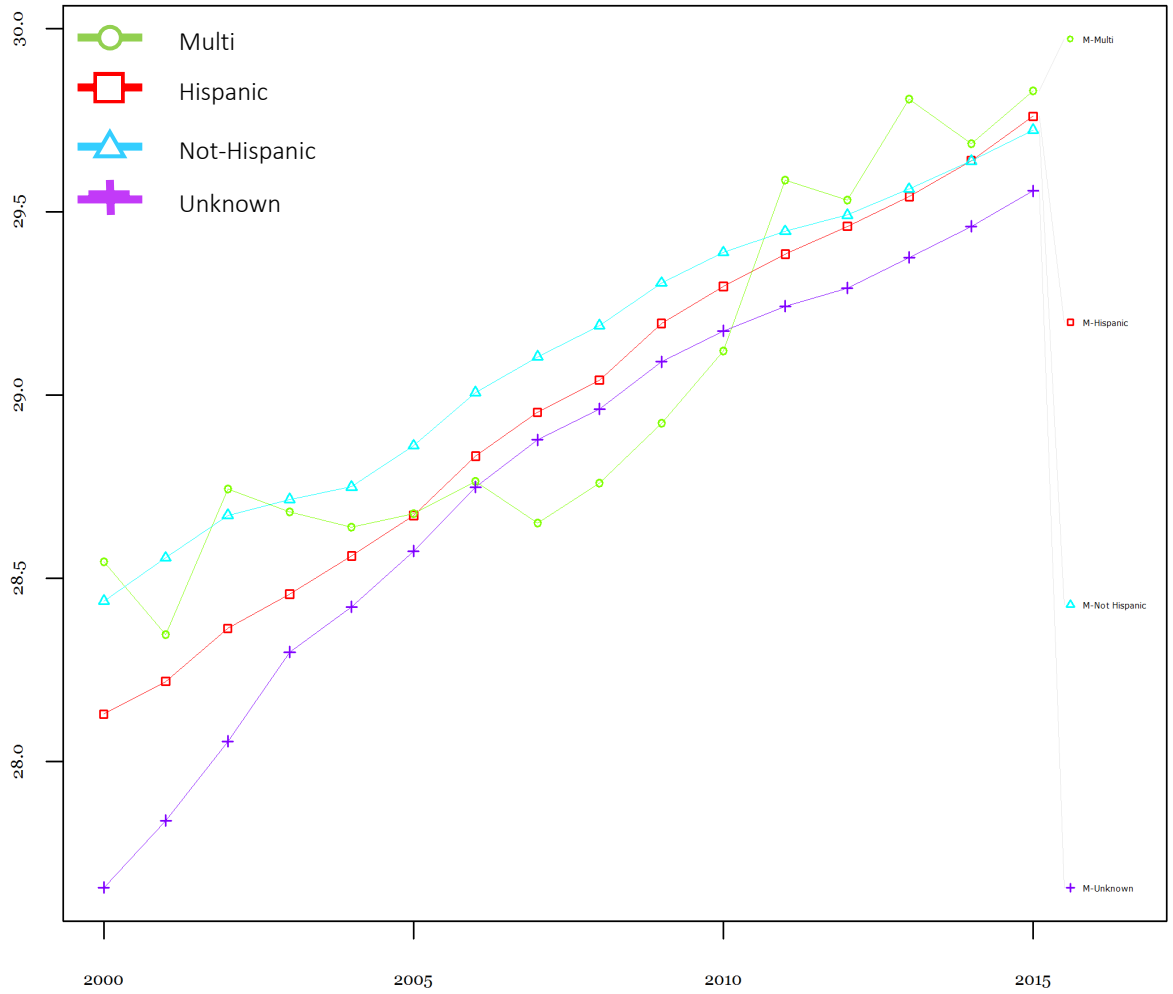
Male

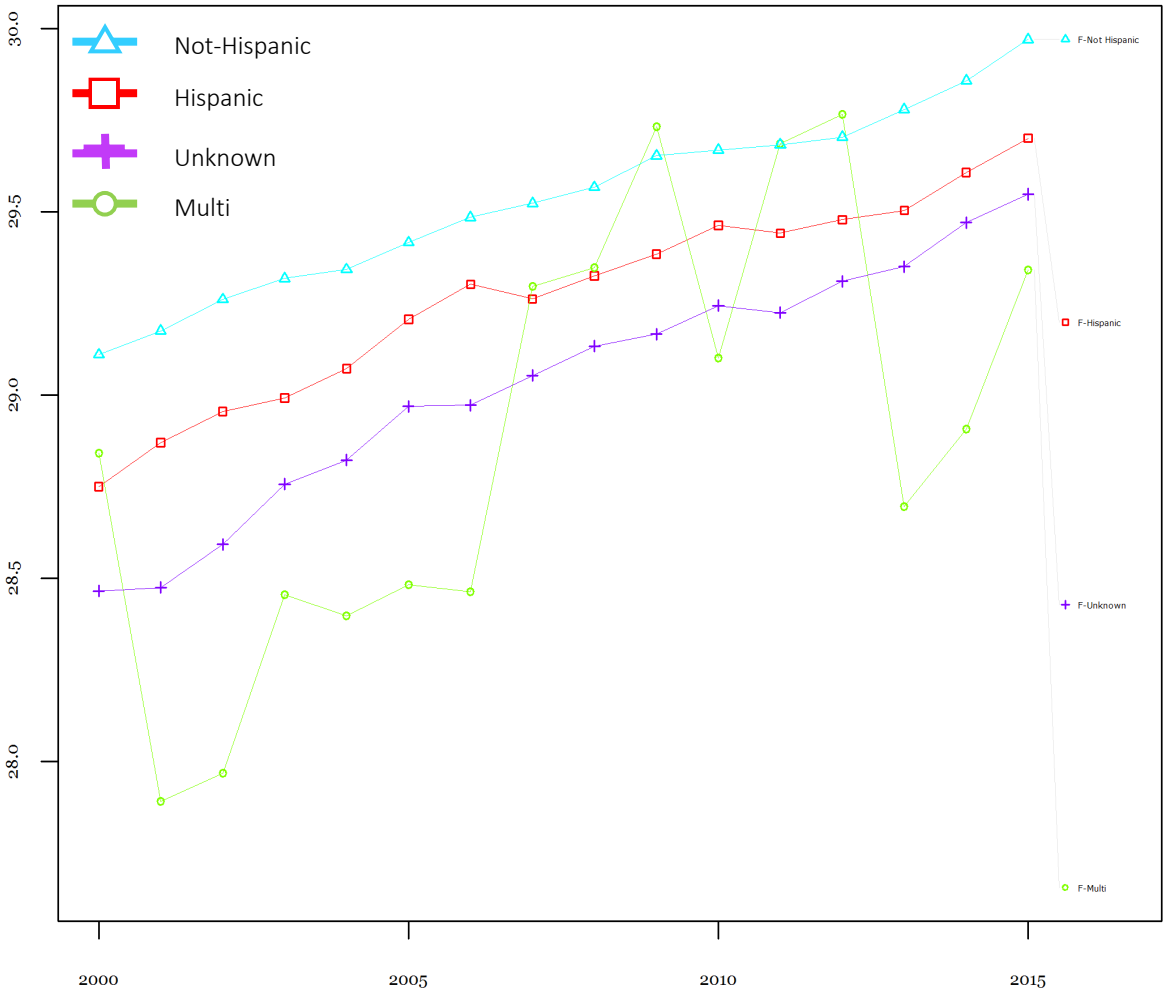


Female

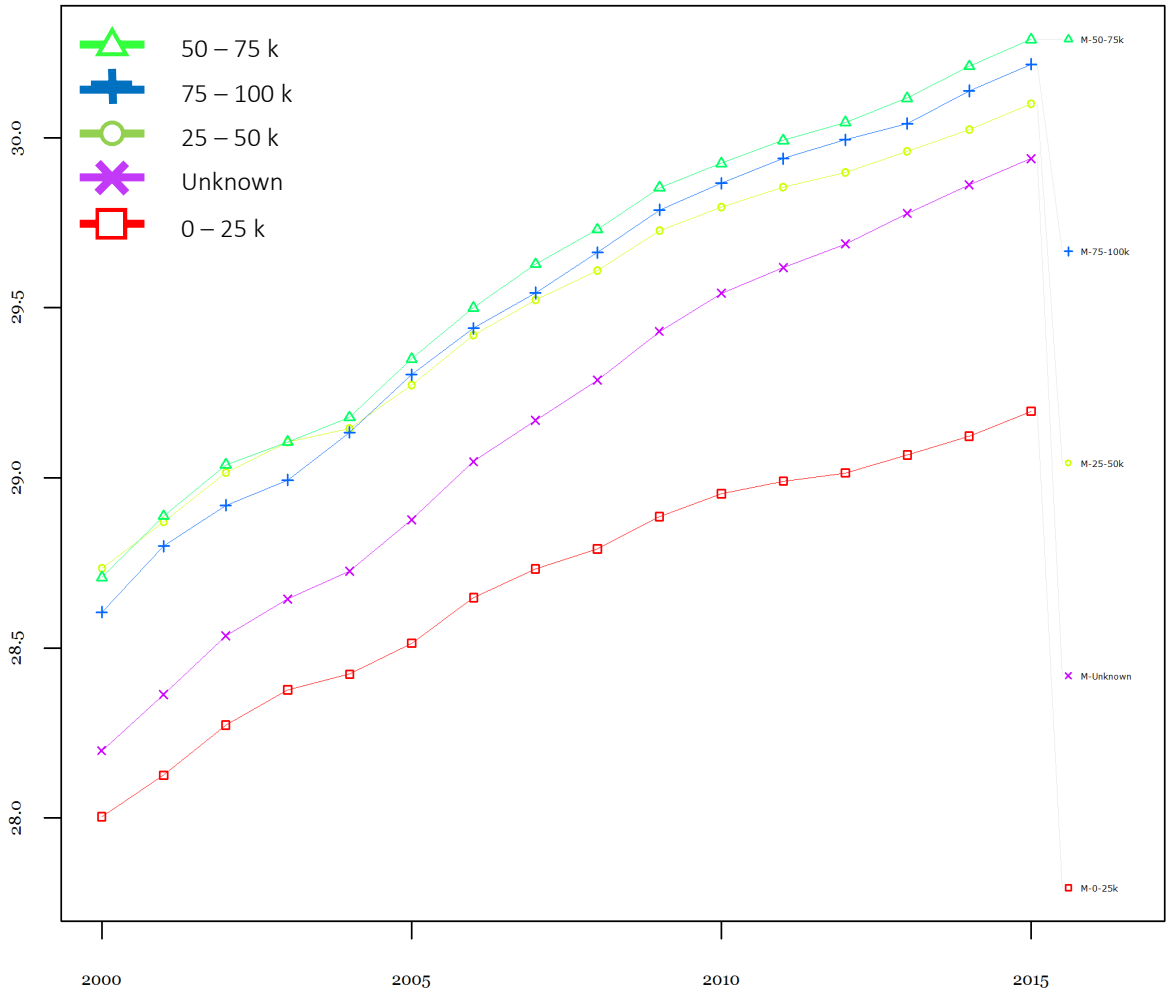


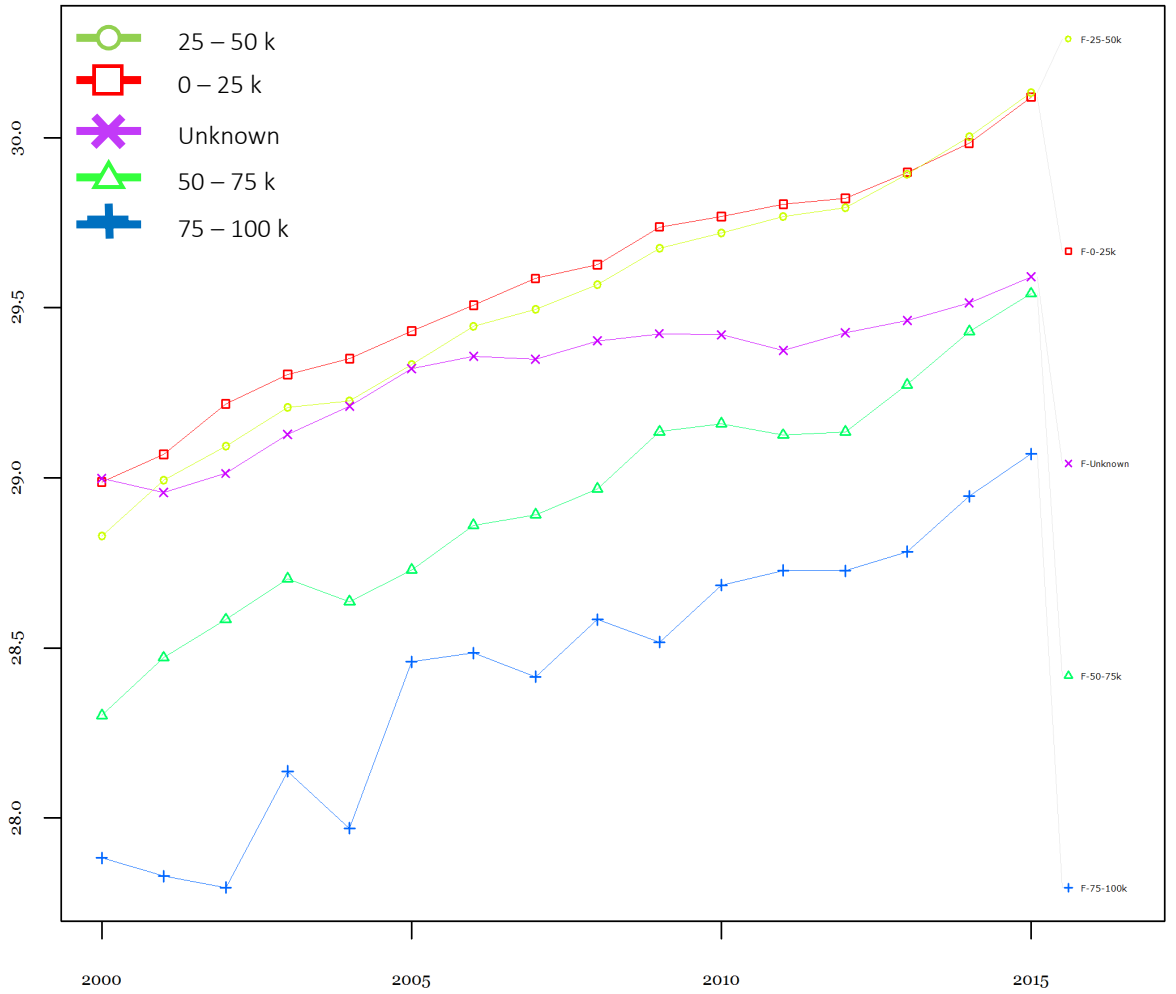
Ethnicity & BMI [LMR] 2000-2015



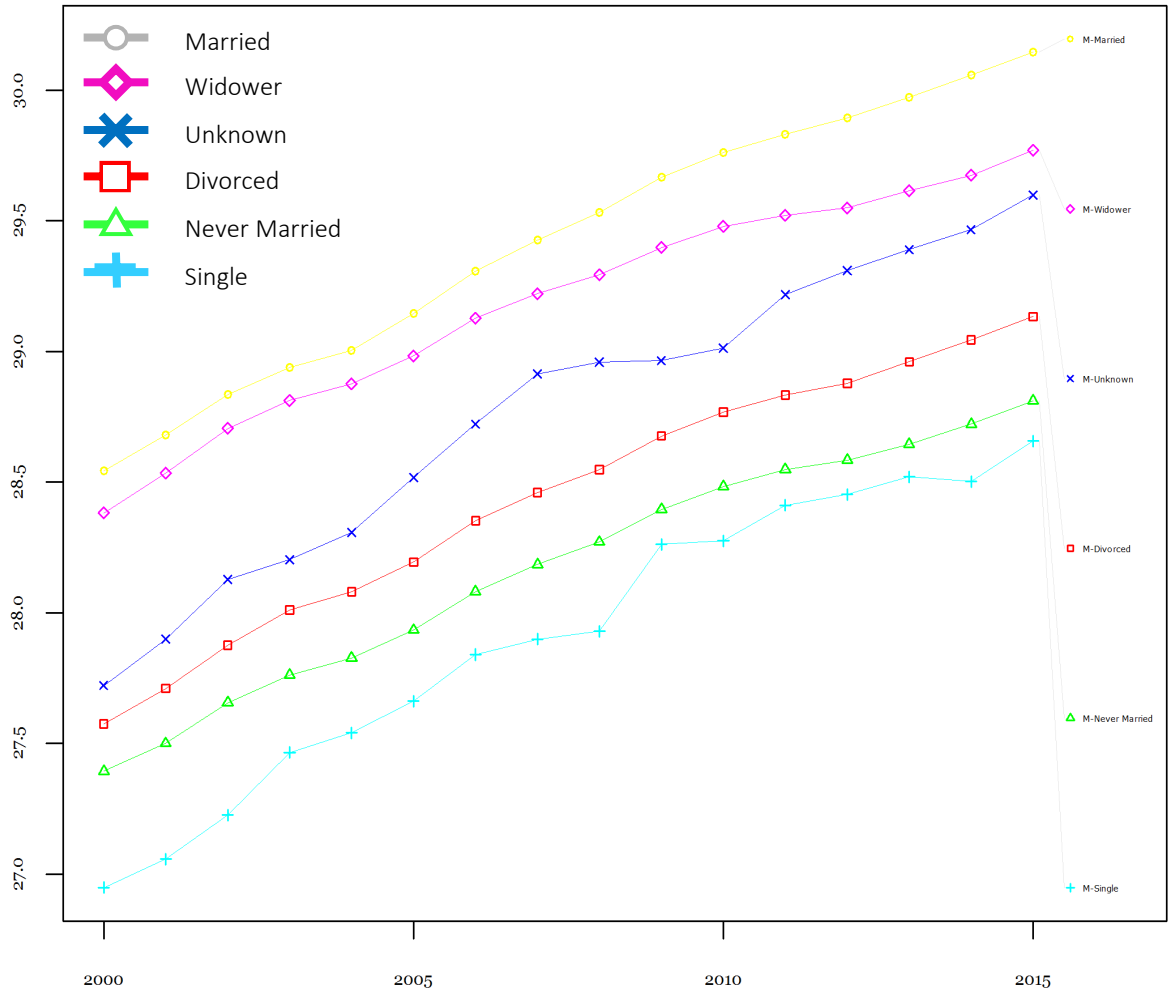


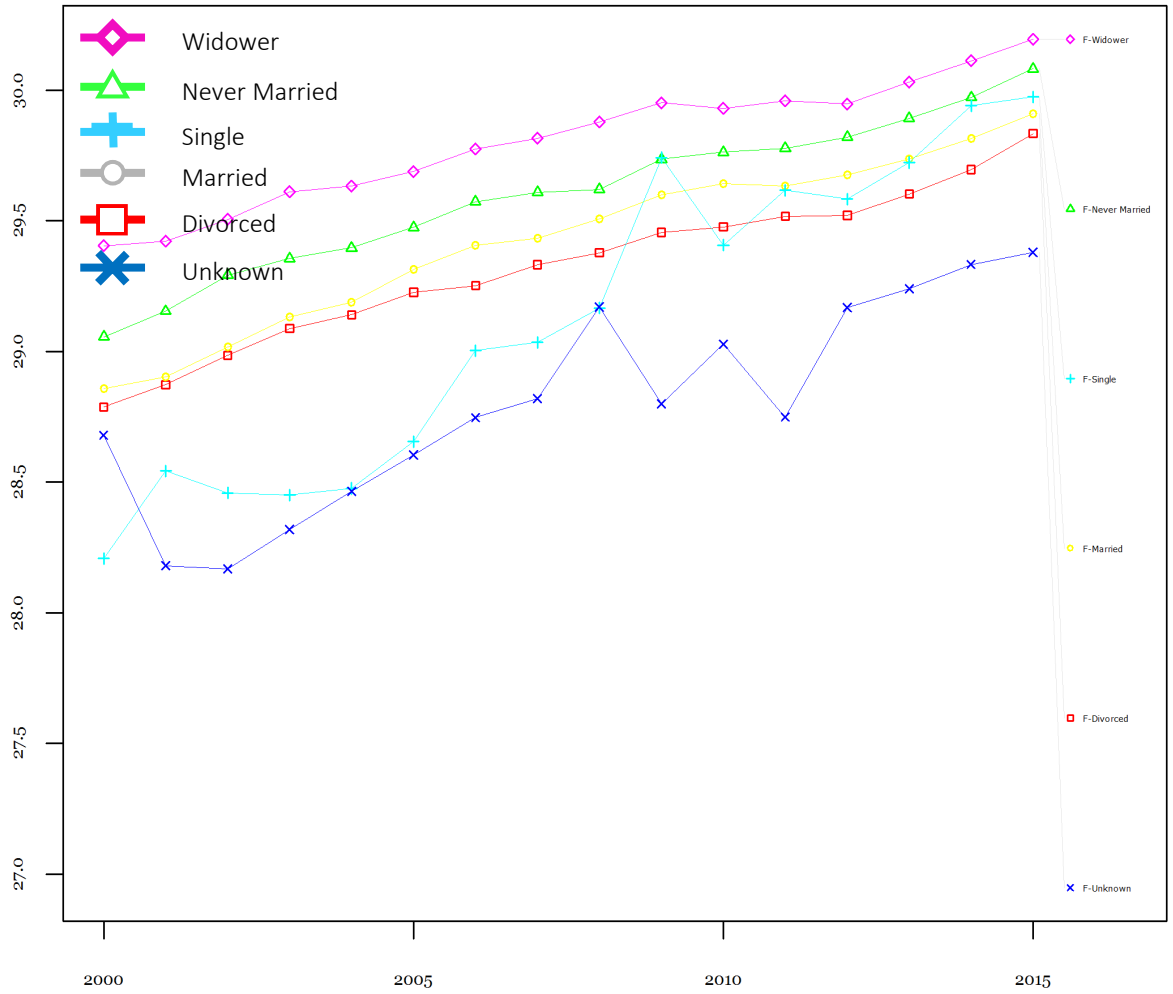
Self-Reported Income & BMI [LMR] 2000-2015





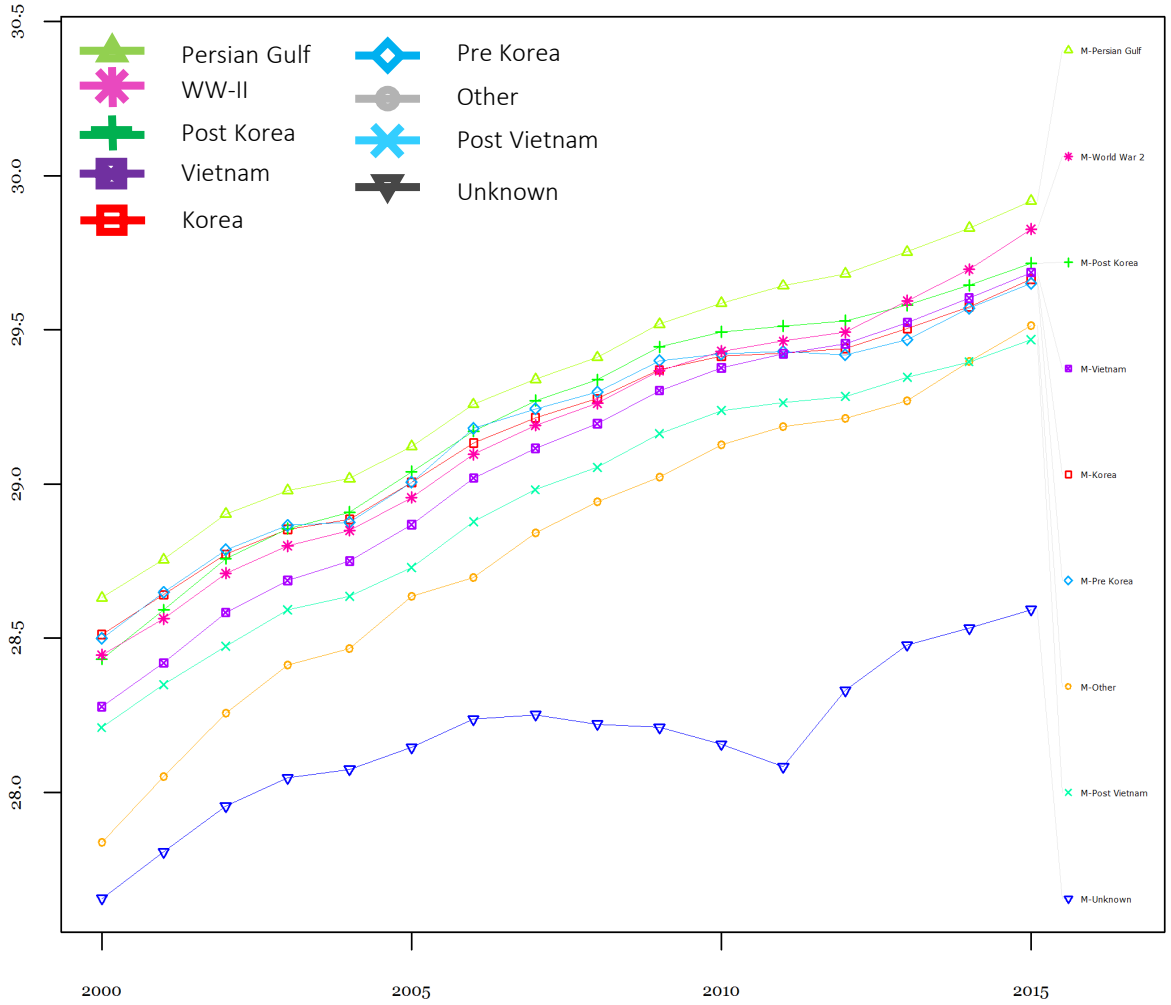
Marital & BMI [LMR] 2000-2015



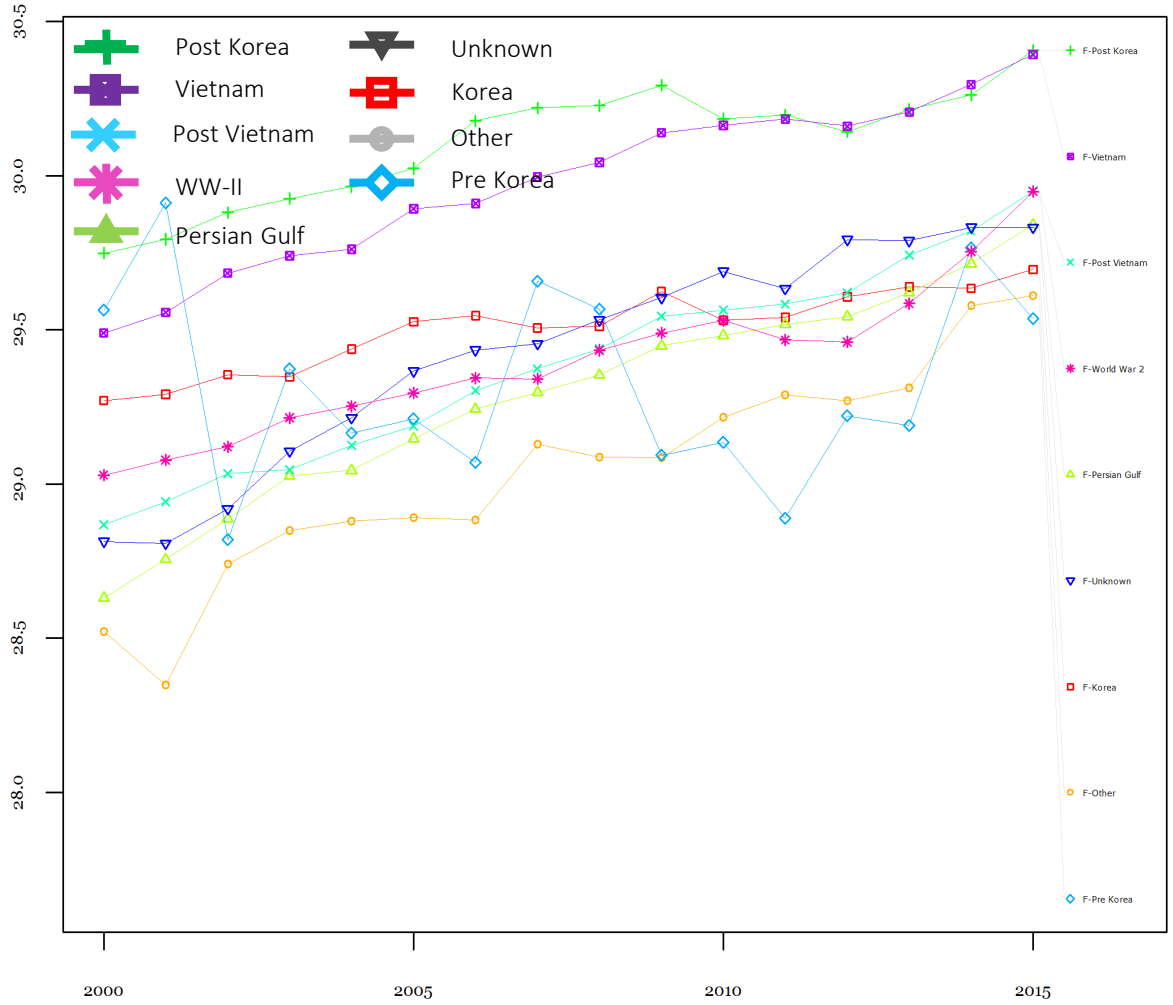


Service & BMI [LMR] 2000-2015

Male

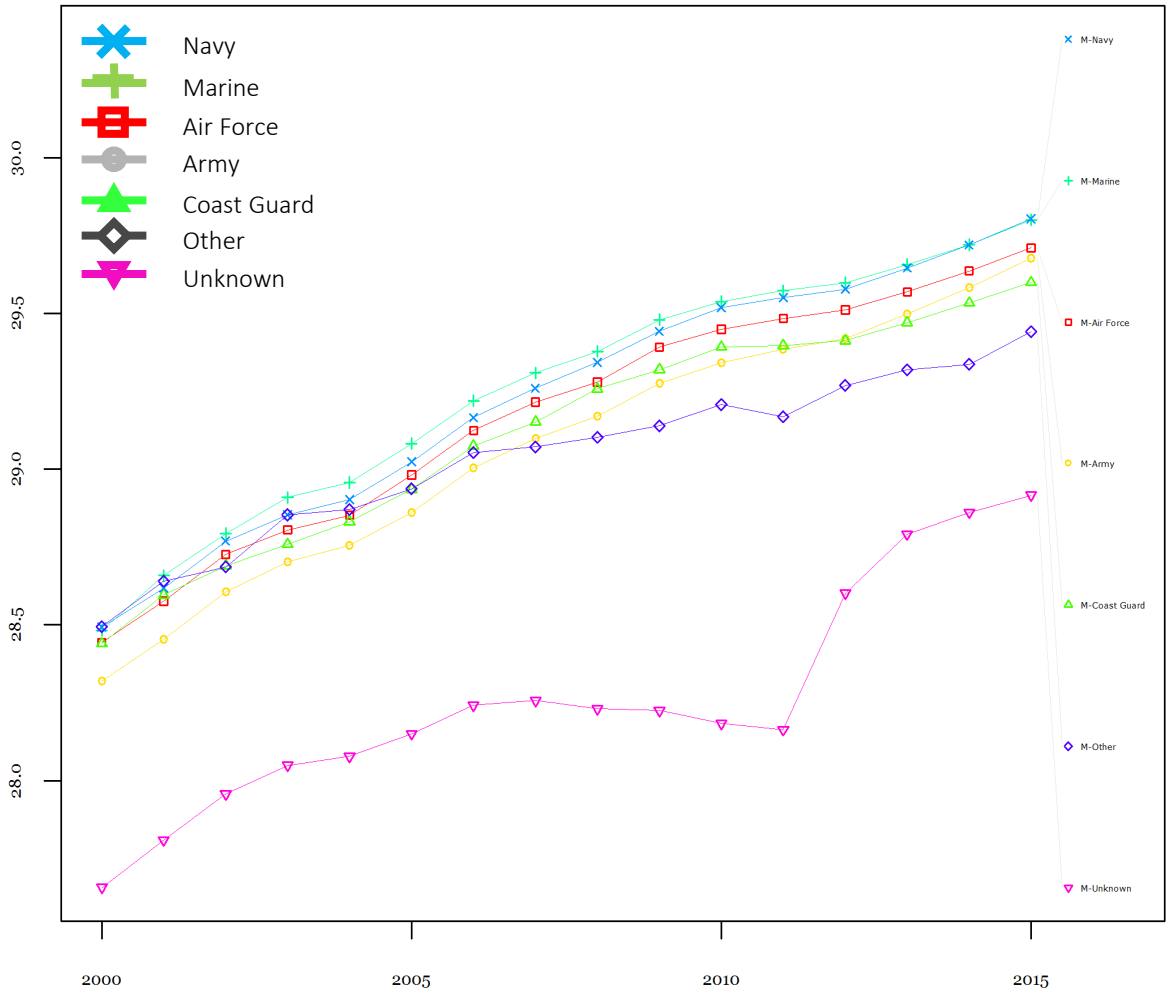


Female

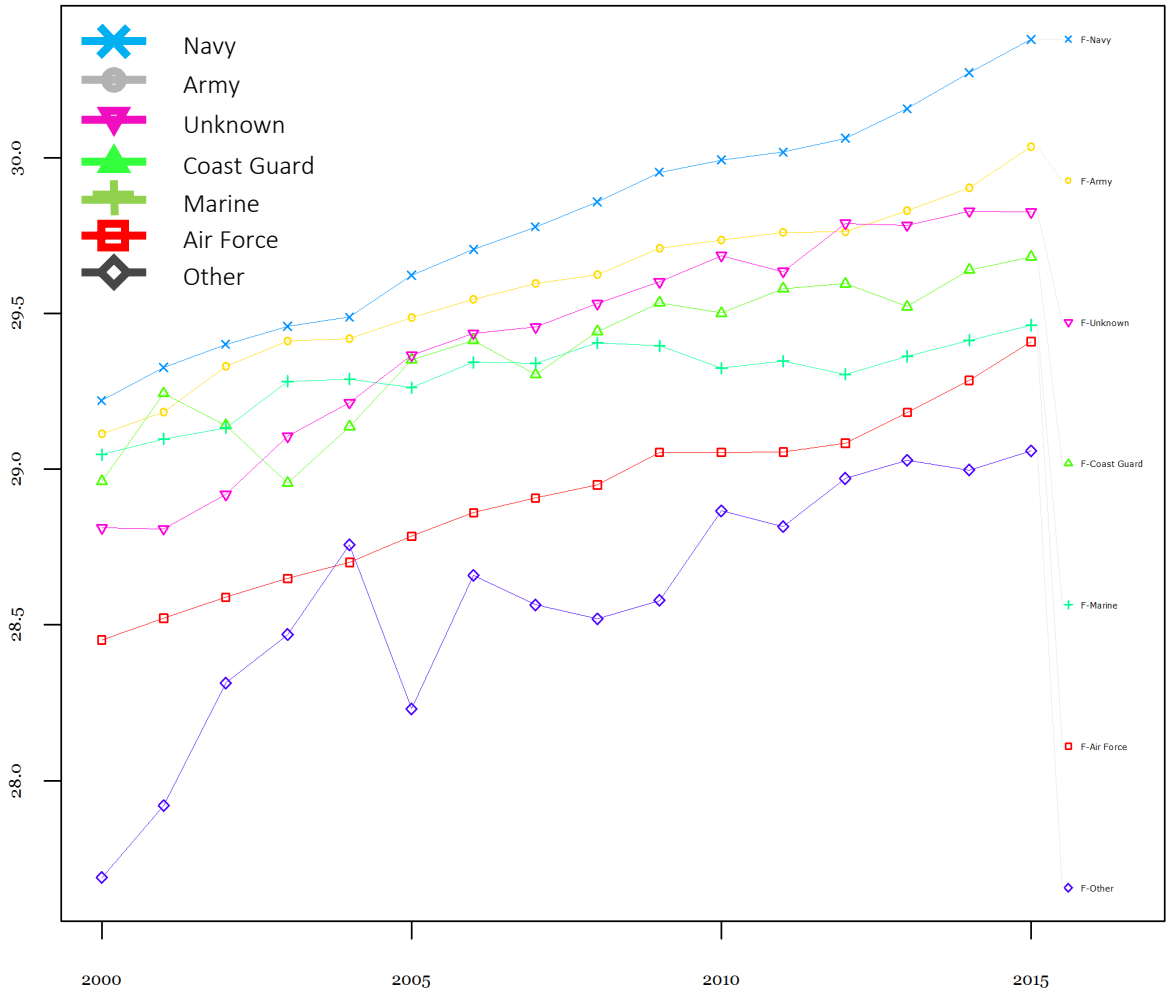


Branch & BMI [LMR] 2000-2015

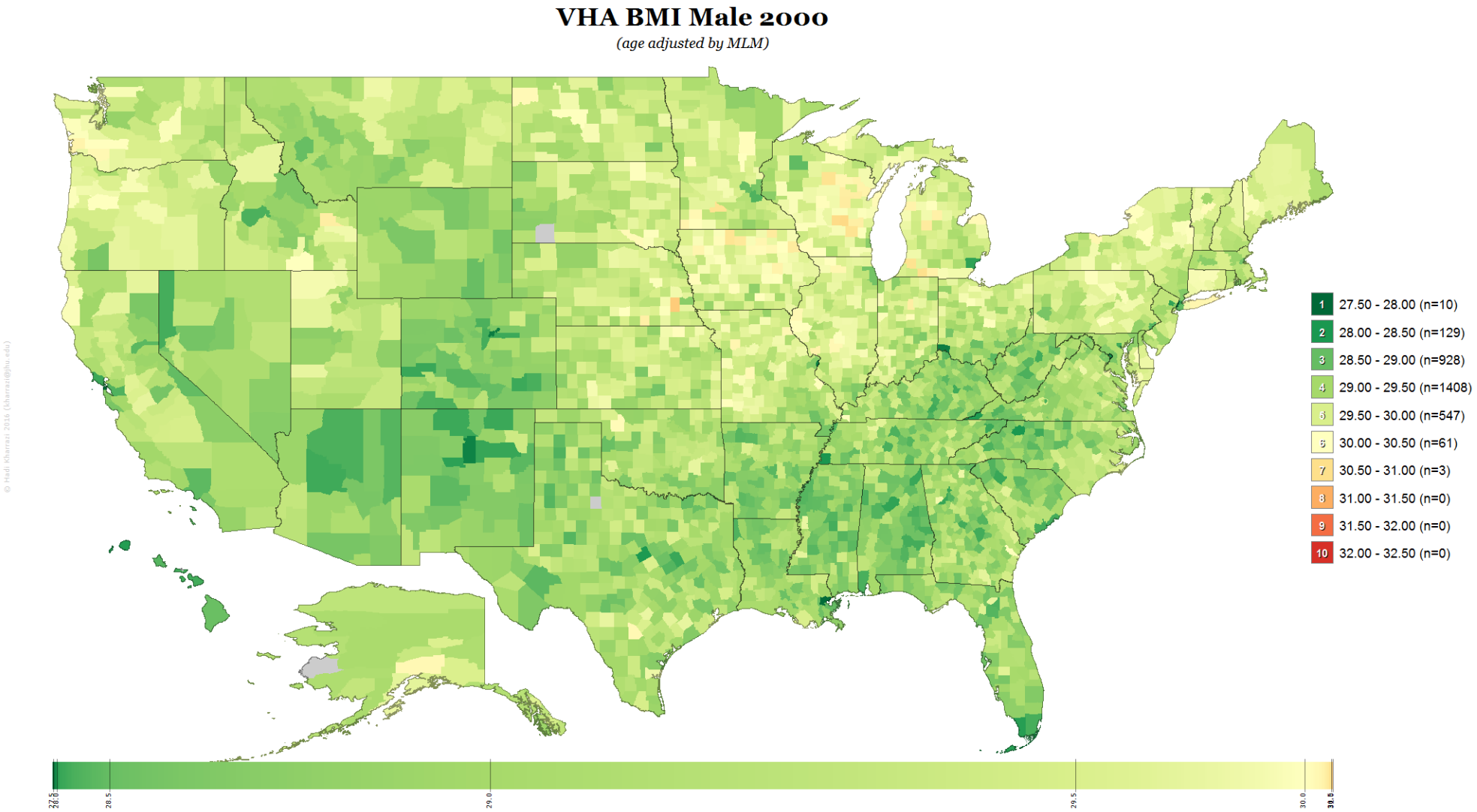
Male



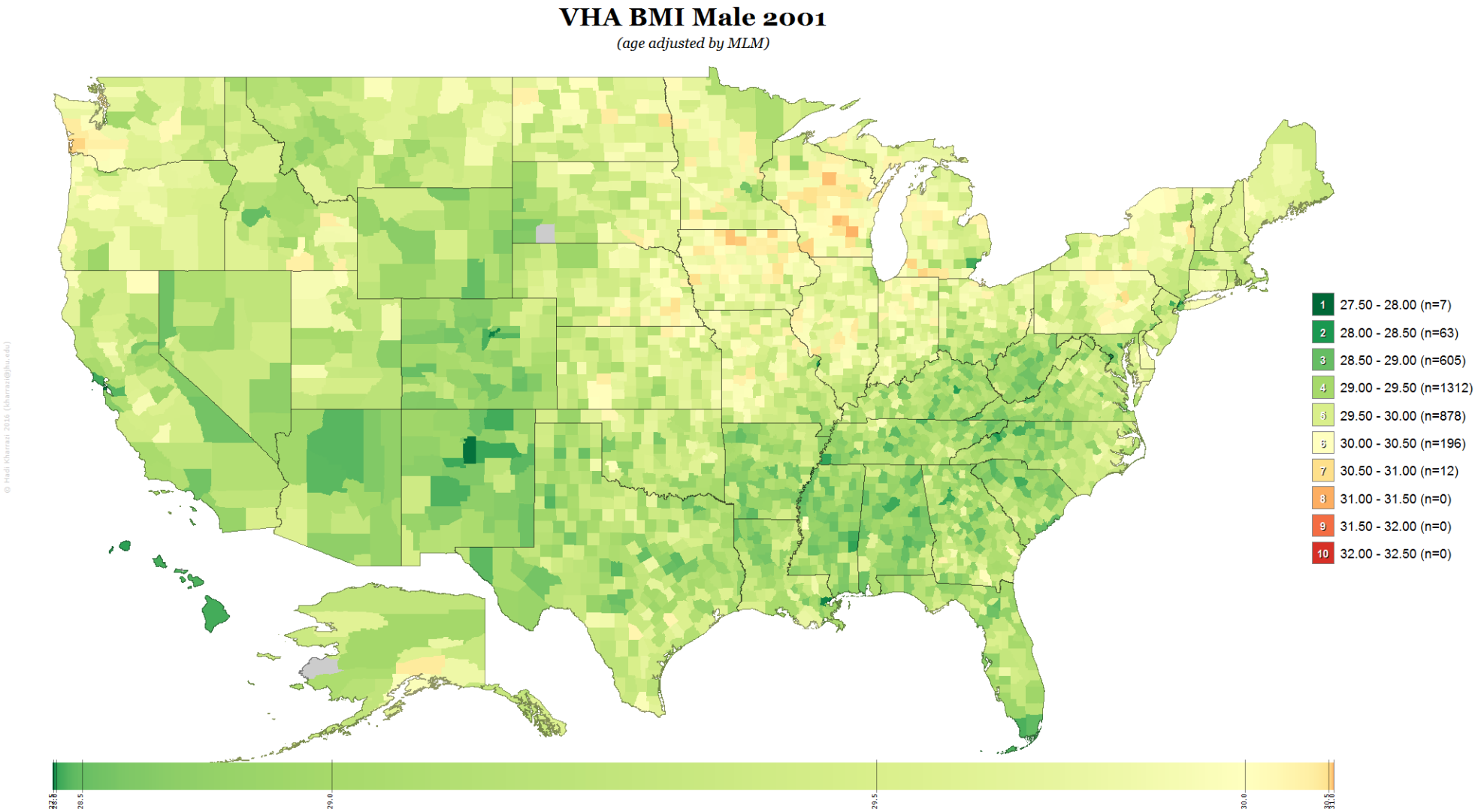
Female



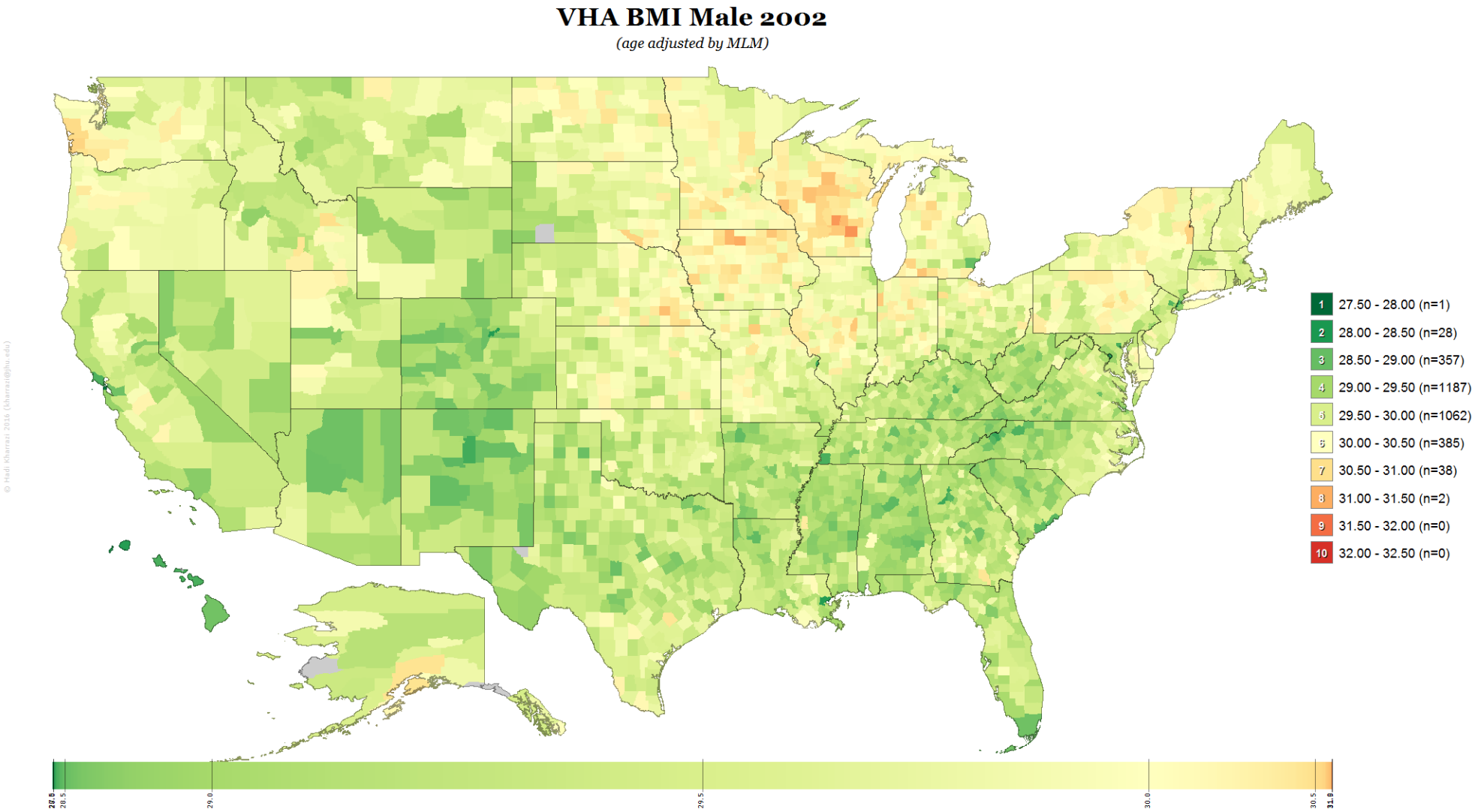
VA Male BMI Age Adjusted [MLM] – 2000



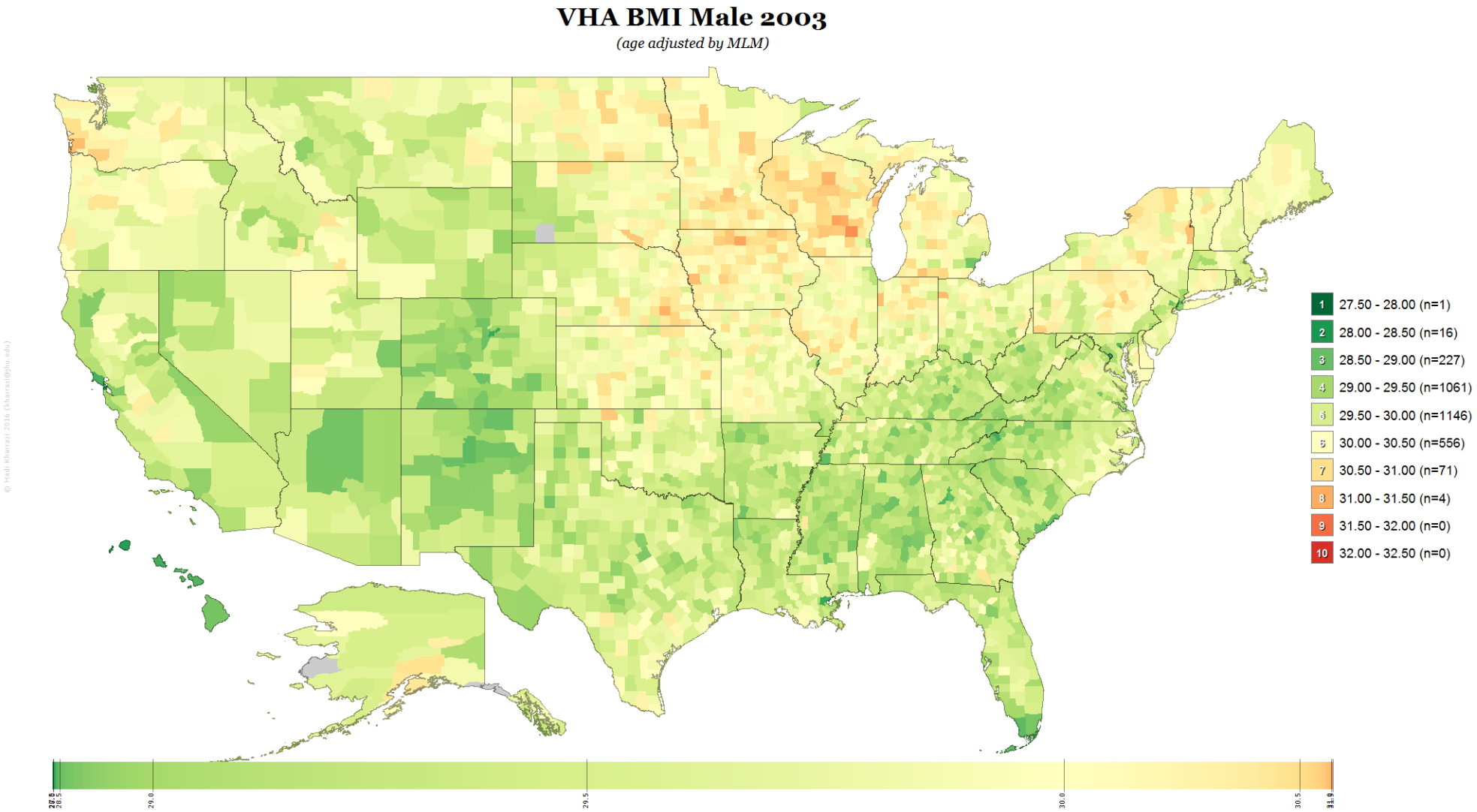
VA Male BMI Age Adjusted [MLM] – 2001



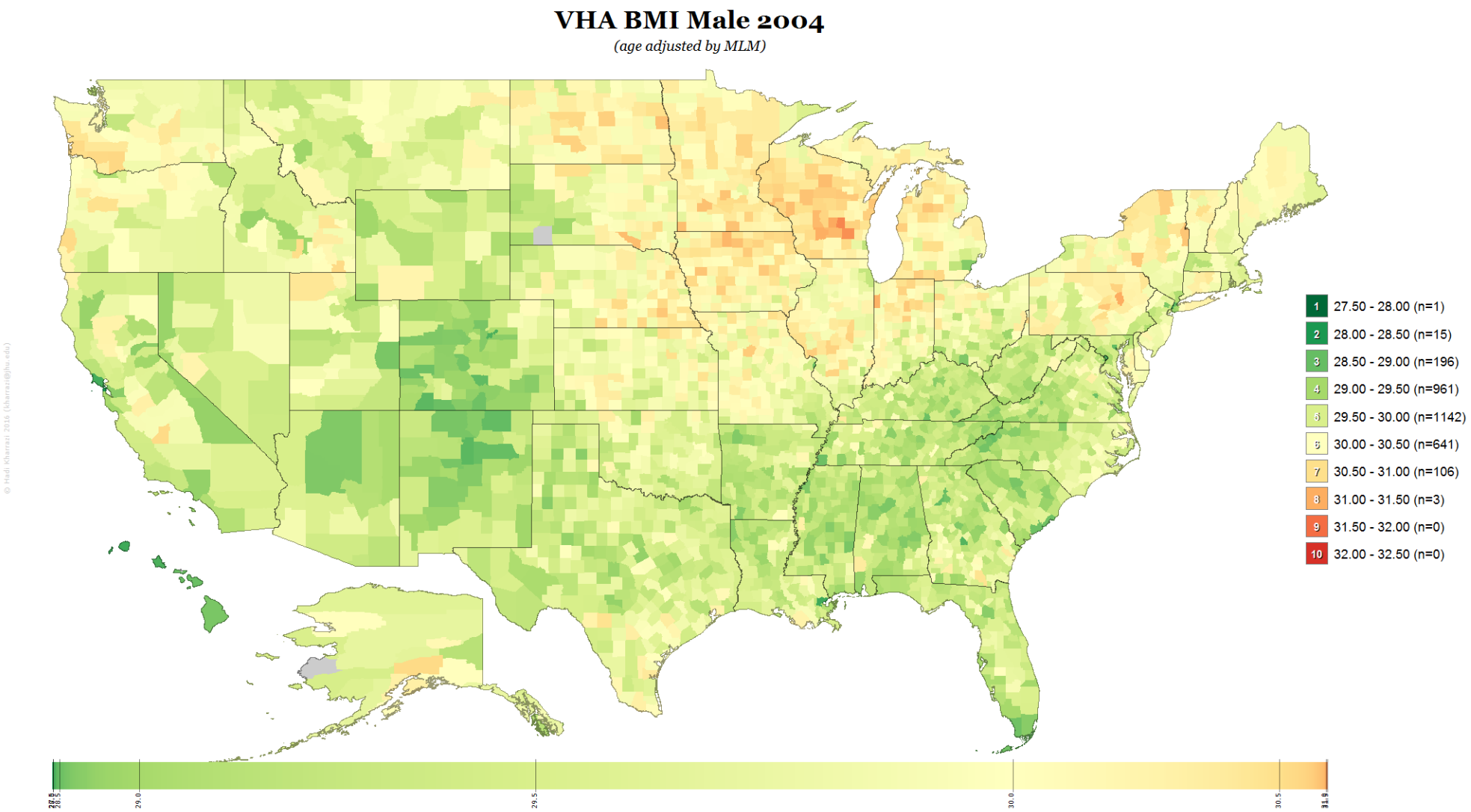
VA Male BMI Age Adjusted [MLM] – 2002



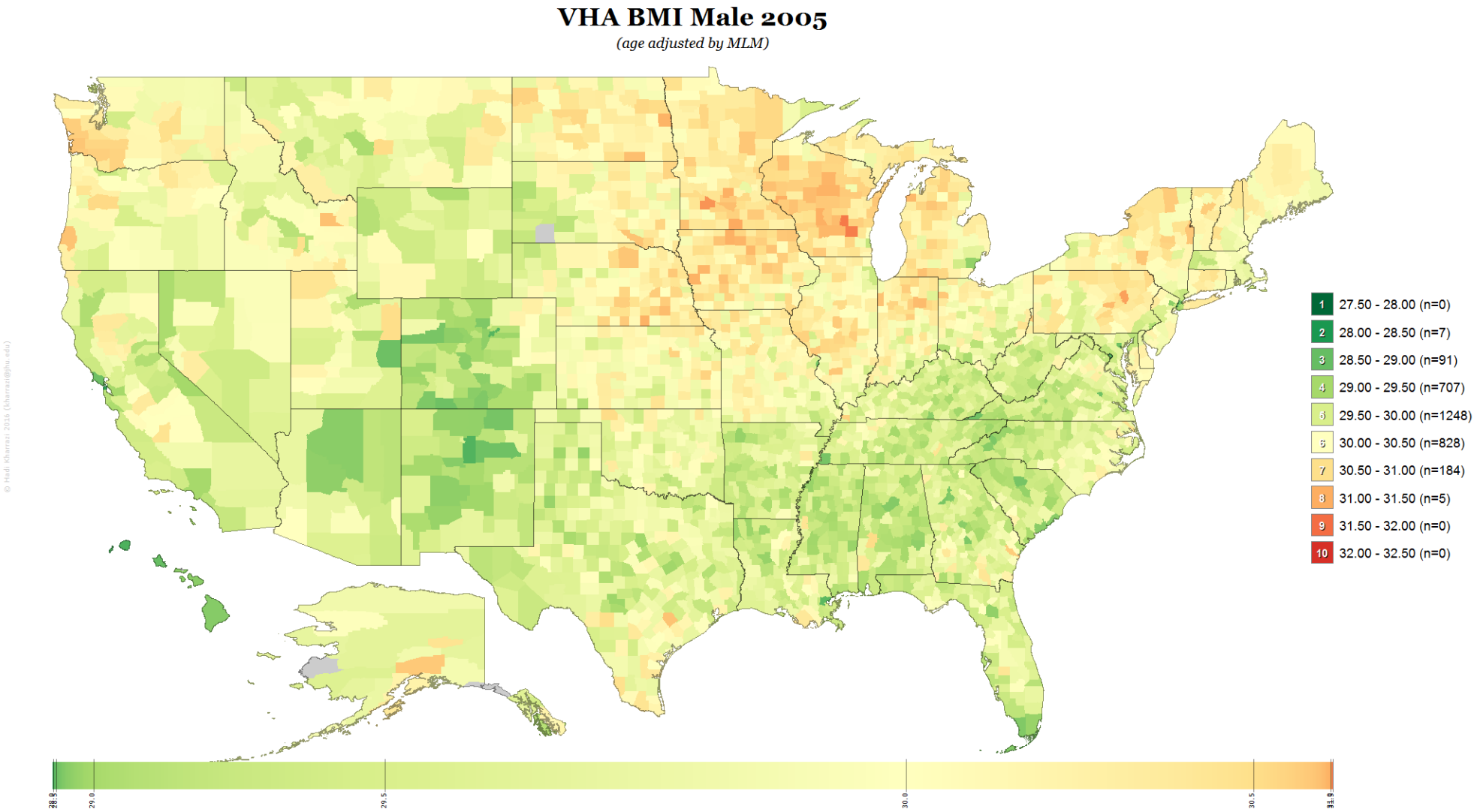
VA Male BMI Age Adjusted [MLM] – 2003



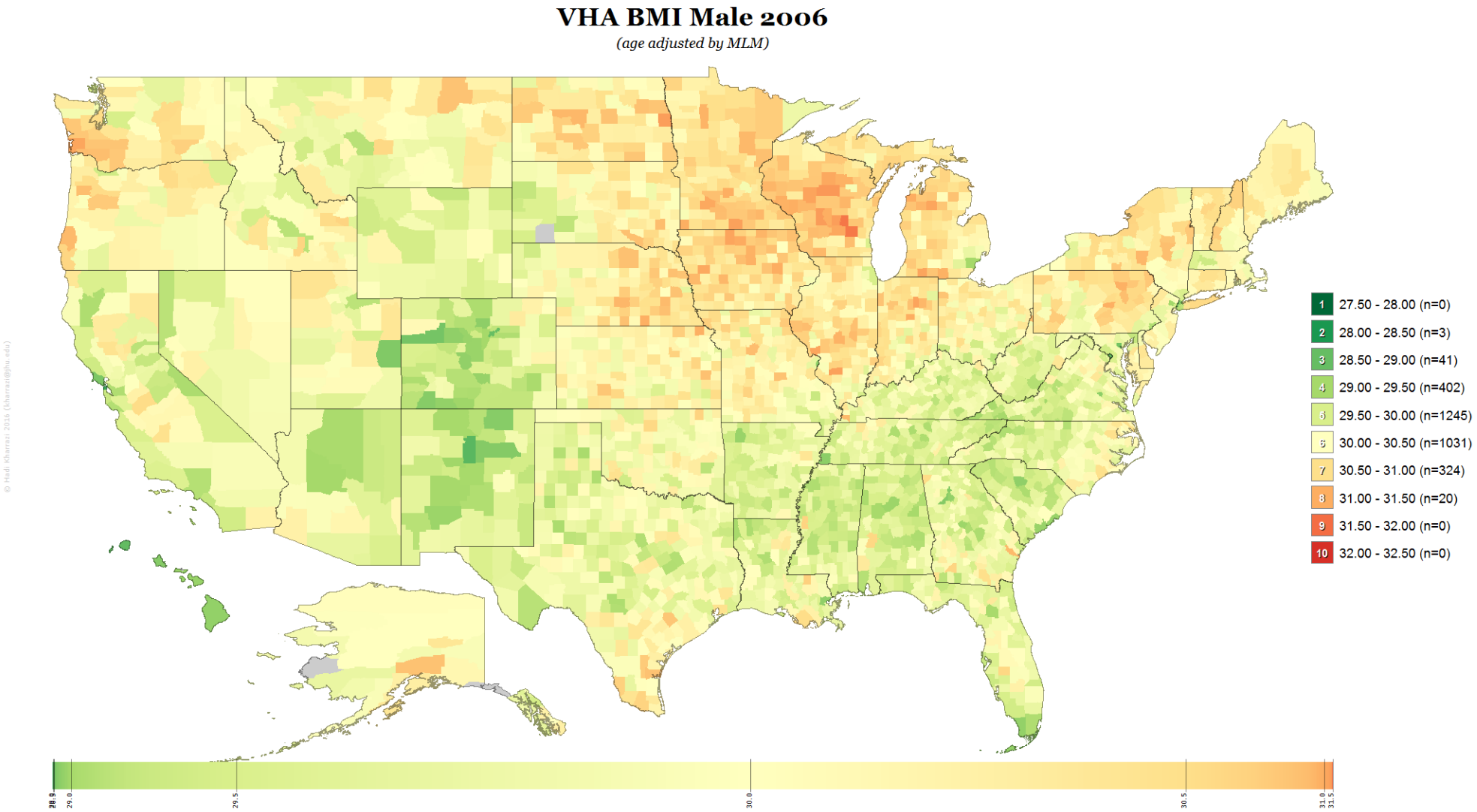
VA Male BMI Age Adjusted [MLM] – 2004



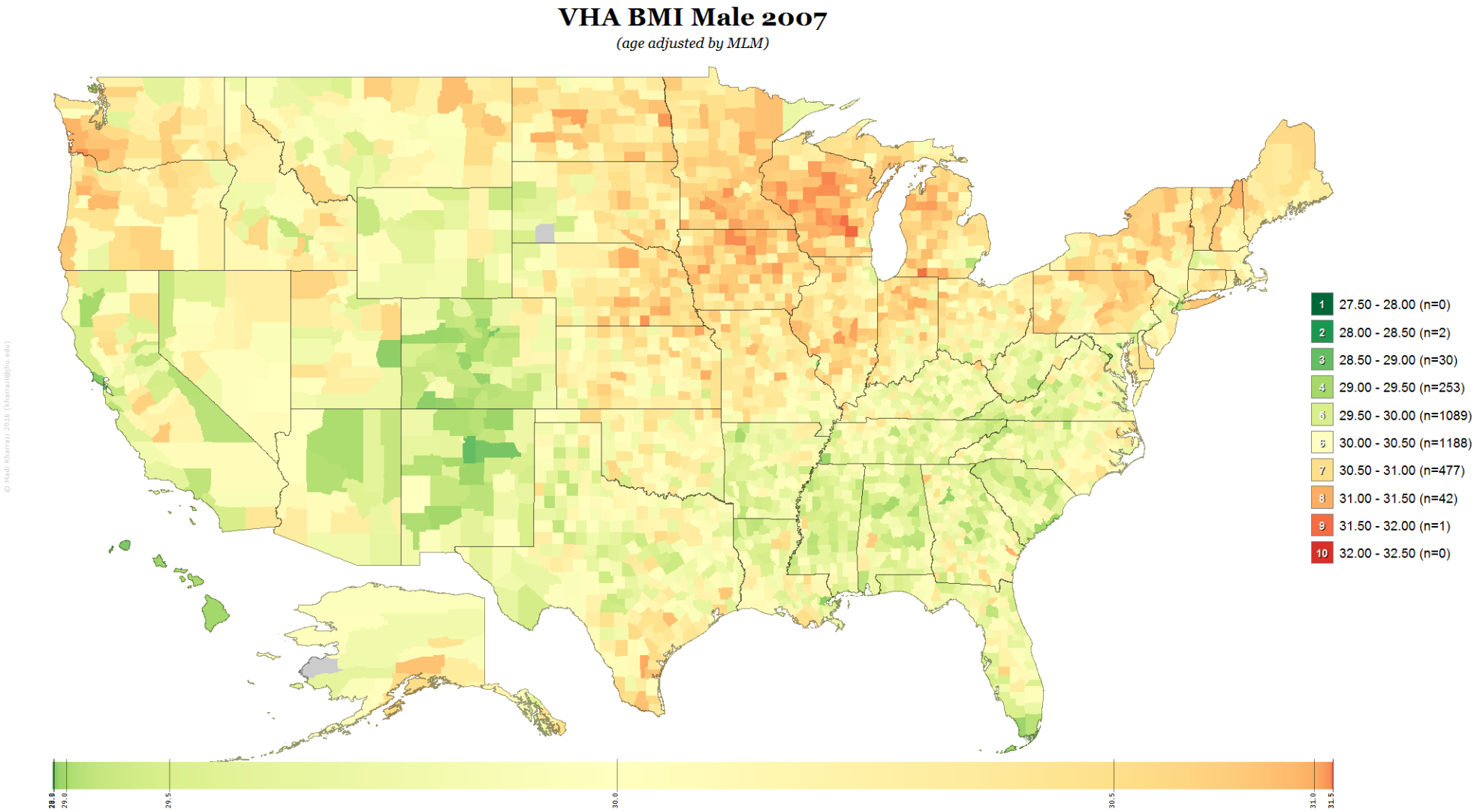
VA Male BMI Age Adjusted [MLM] – 2005



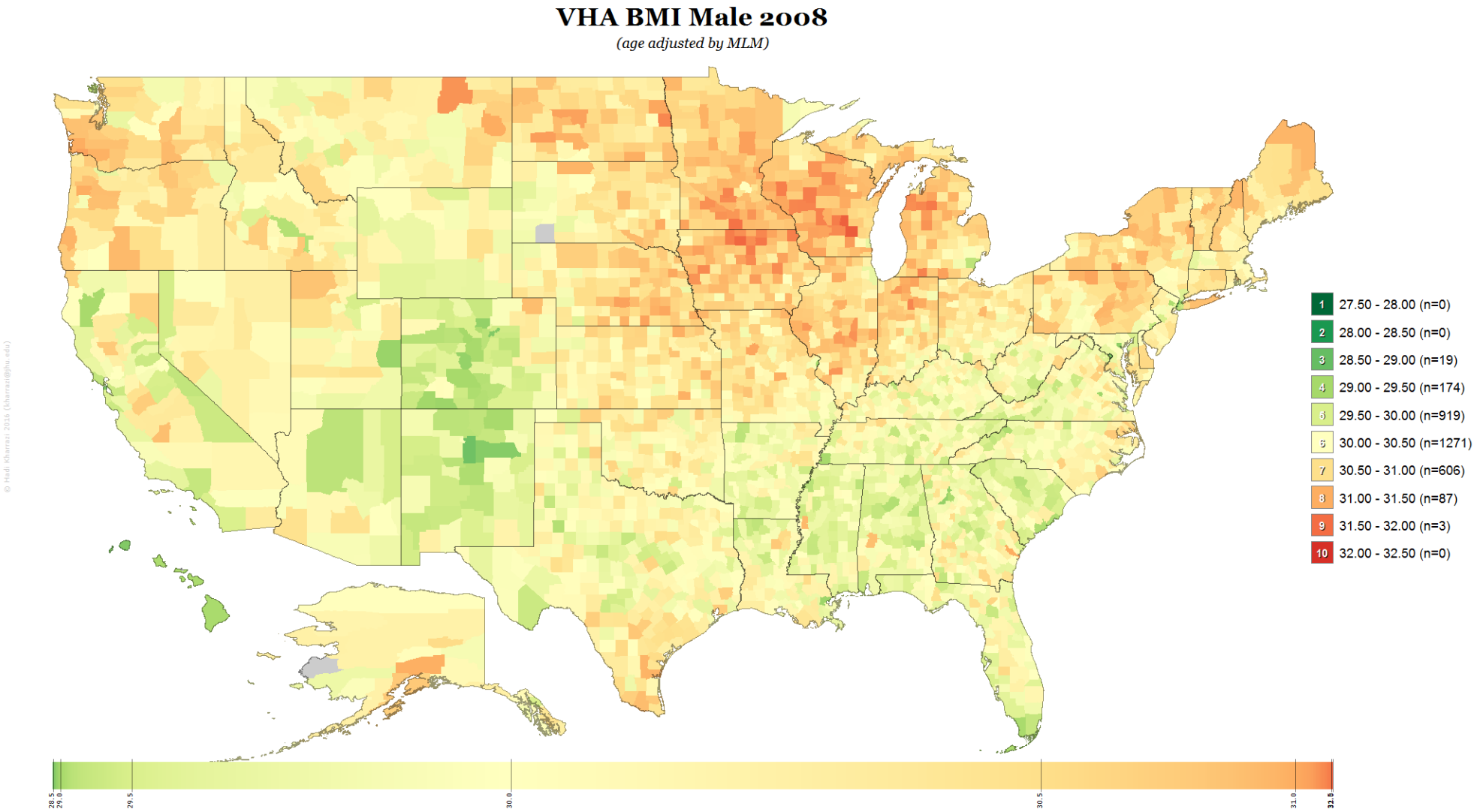
VA Male BMI Age Adjusted [MLM] – 2006



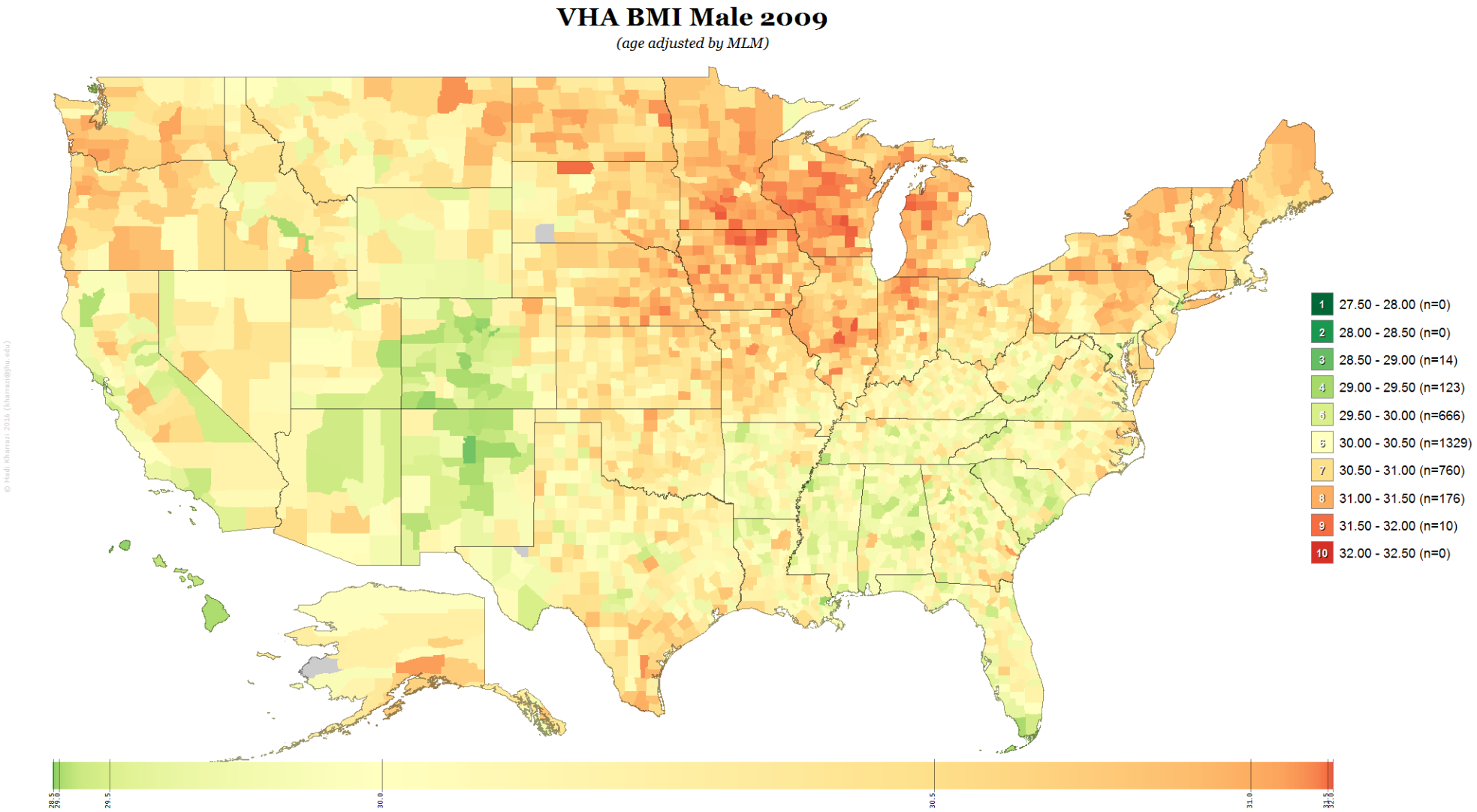
VA Male BMI Age Adjusted [MLM] – 2007



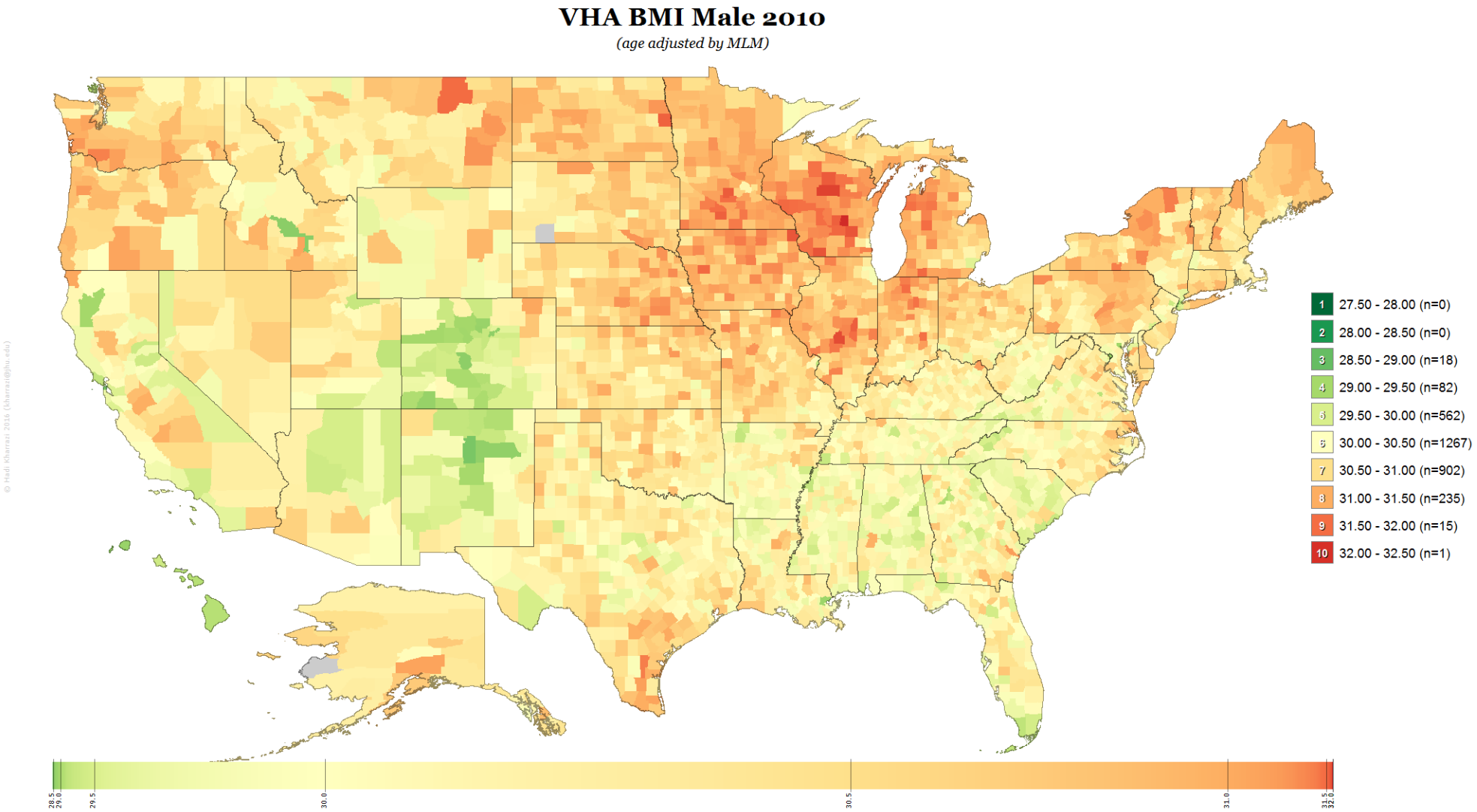
VA Male BMI Age Adjusted [MLM] – 2008



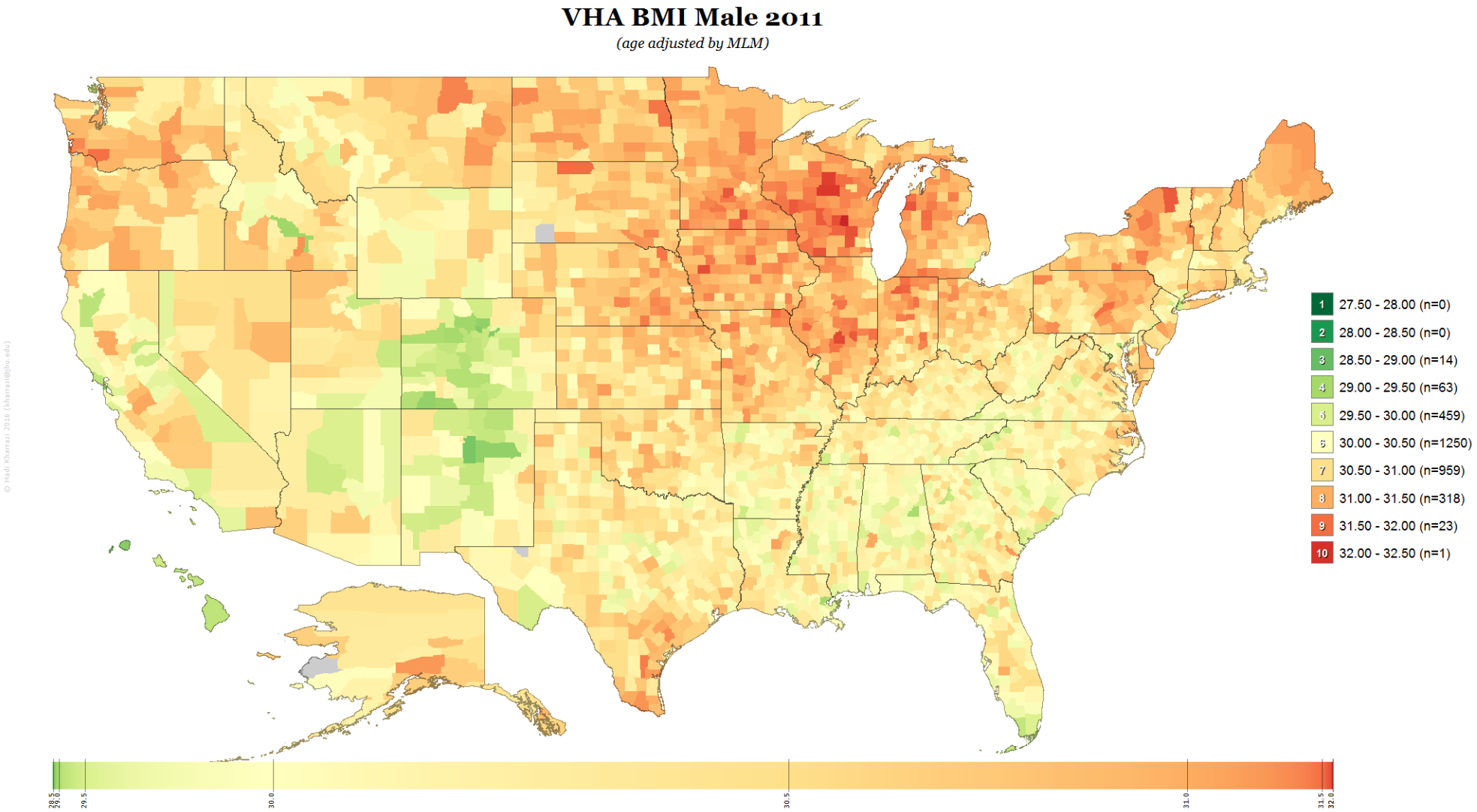
VA Male BMI Age Adjusted [MLM] – 2009



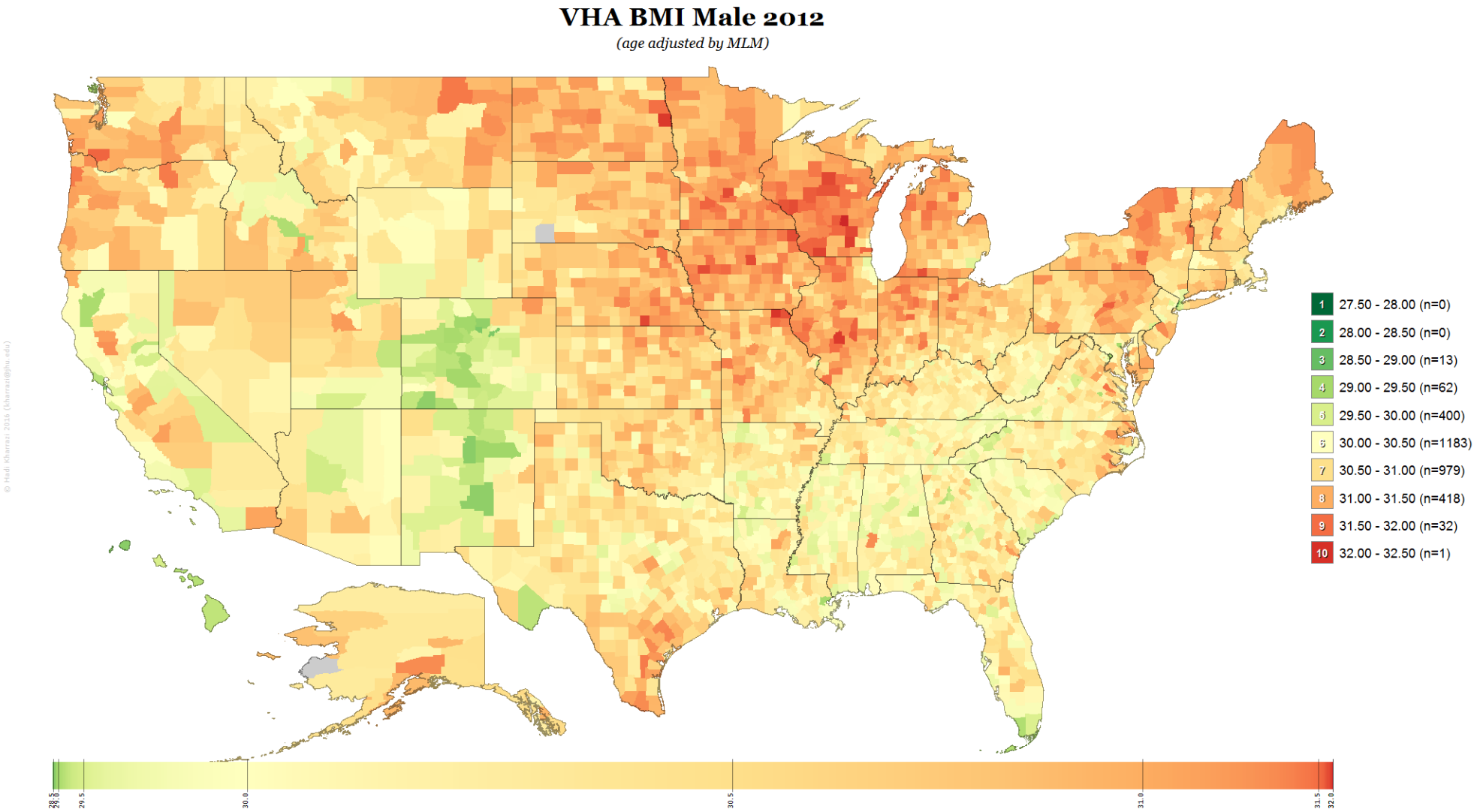
VA Male BMI Age Adjusted [MLM] – 2010



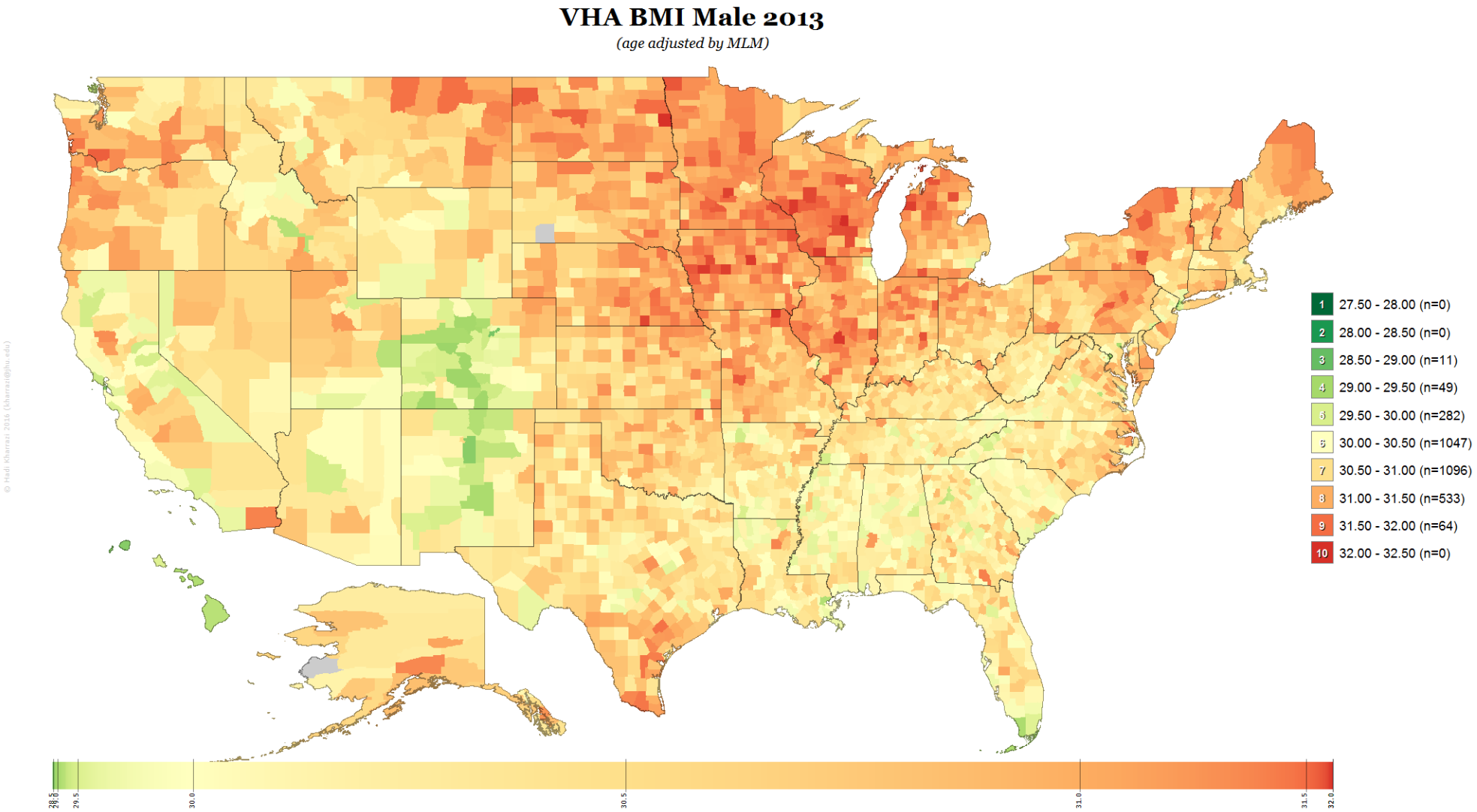
VA Male BMI Age Adjusted [MLM] – 2011



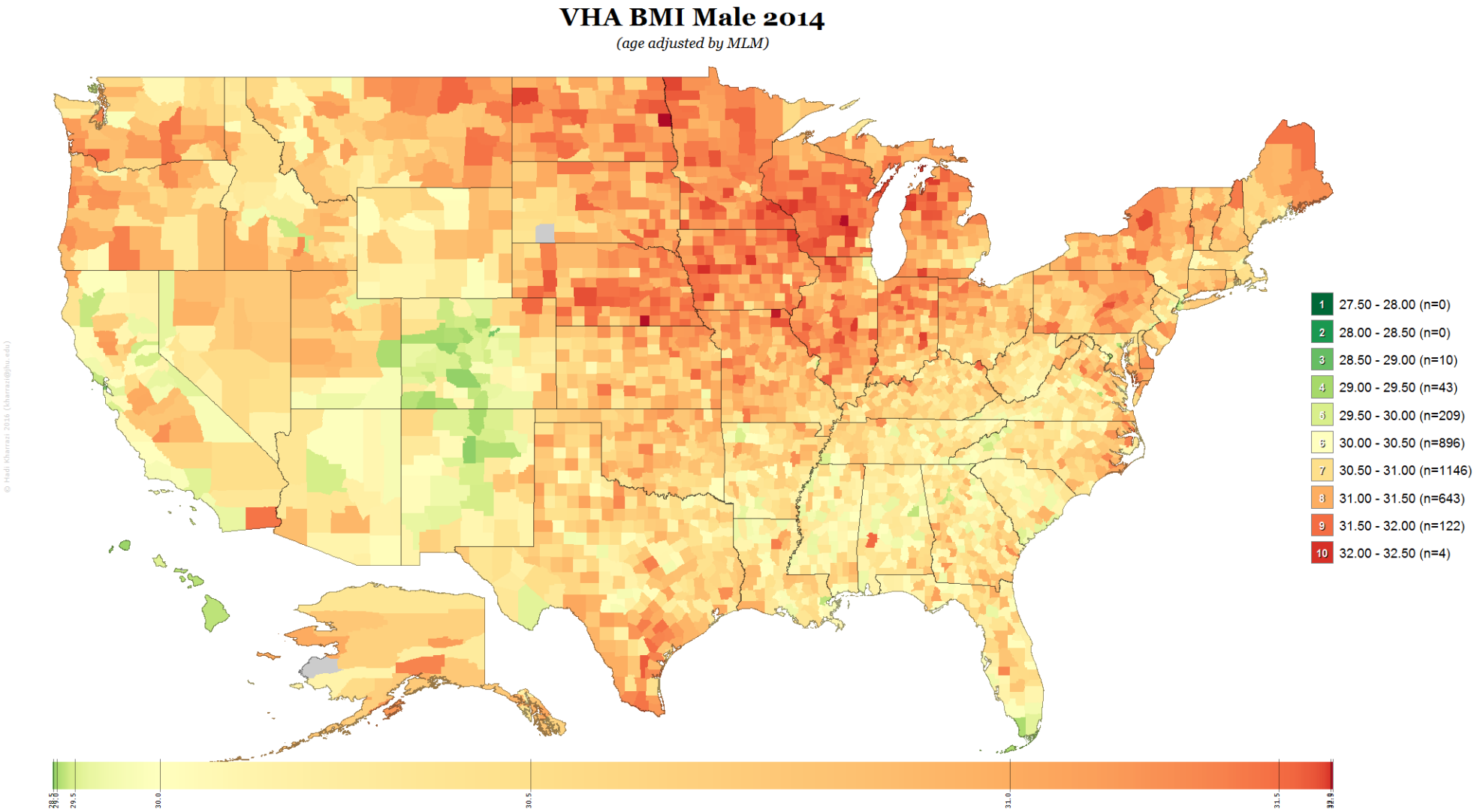
VA Male BMI Age Adjusted [MLM] – 2012



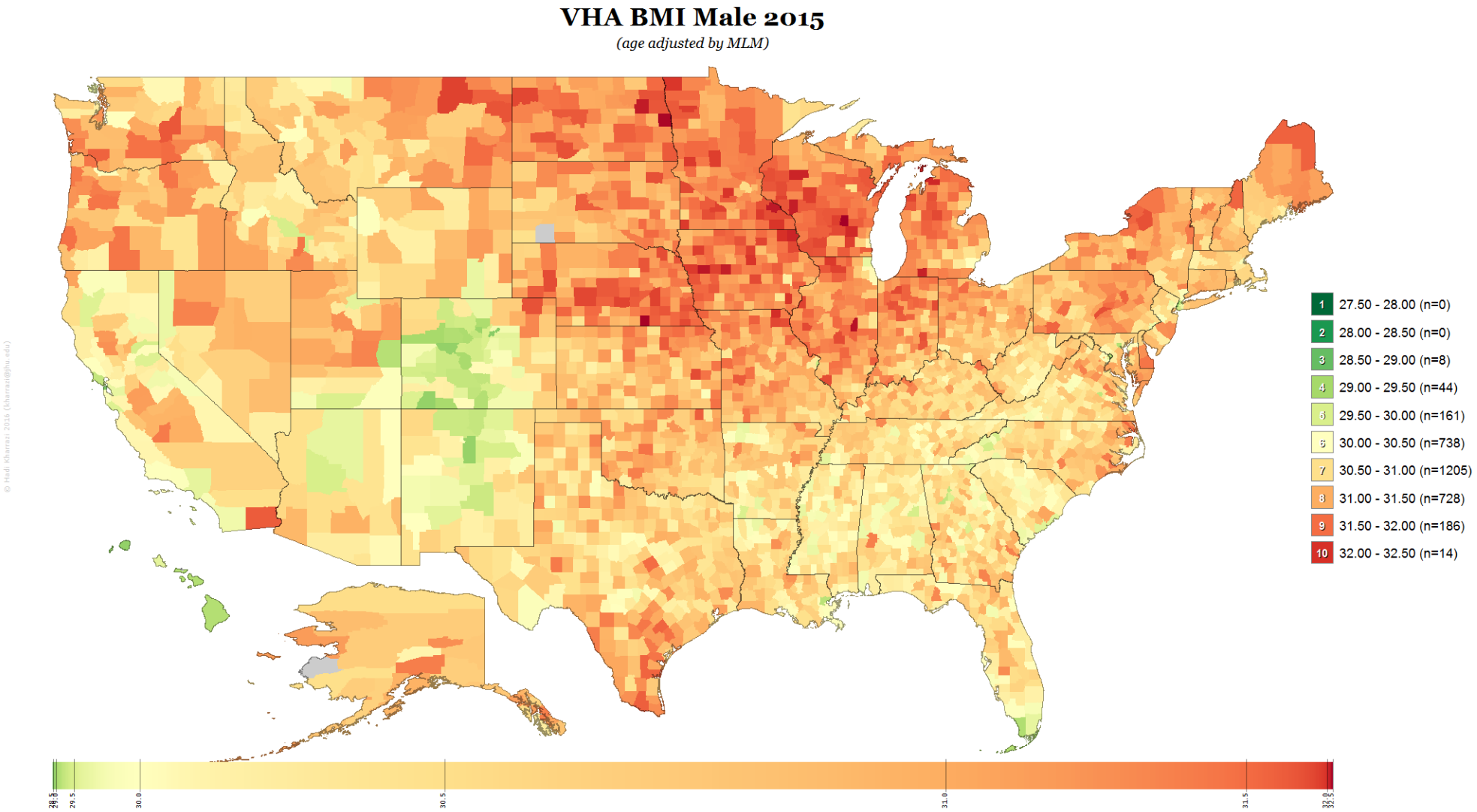
VA Male BMI Age Adjusted [MLM] – 2013



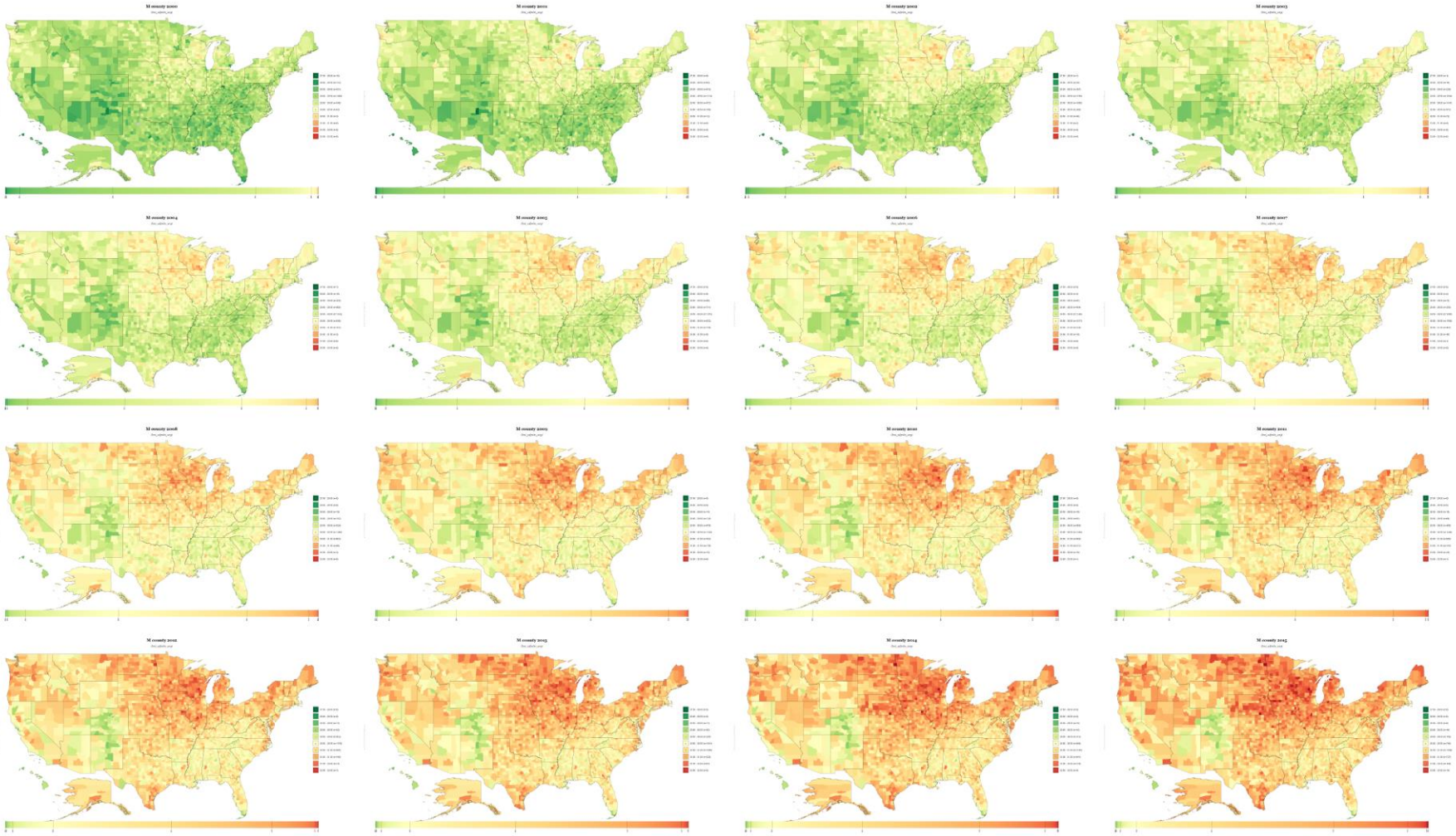
VA Male BMI Age Adjusted [MLM] – 2014



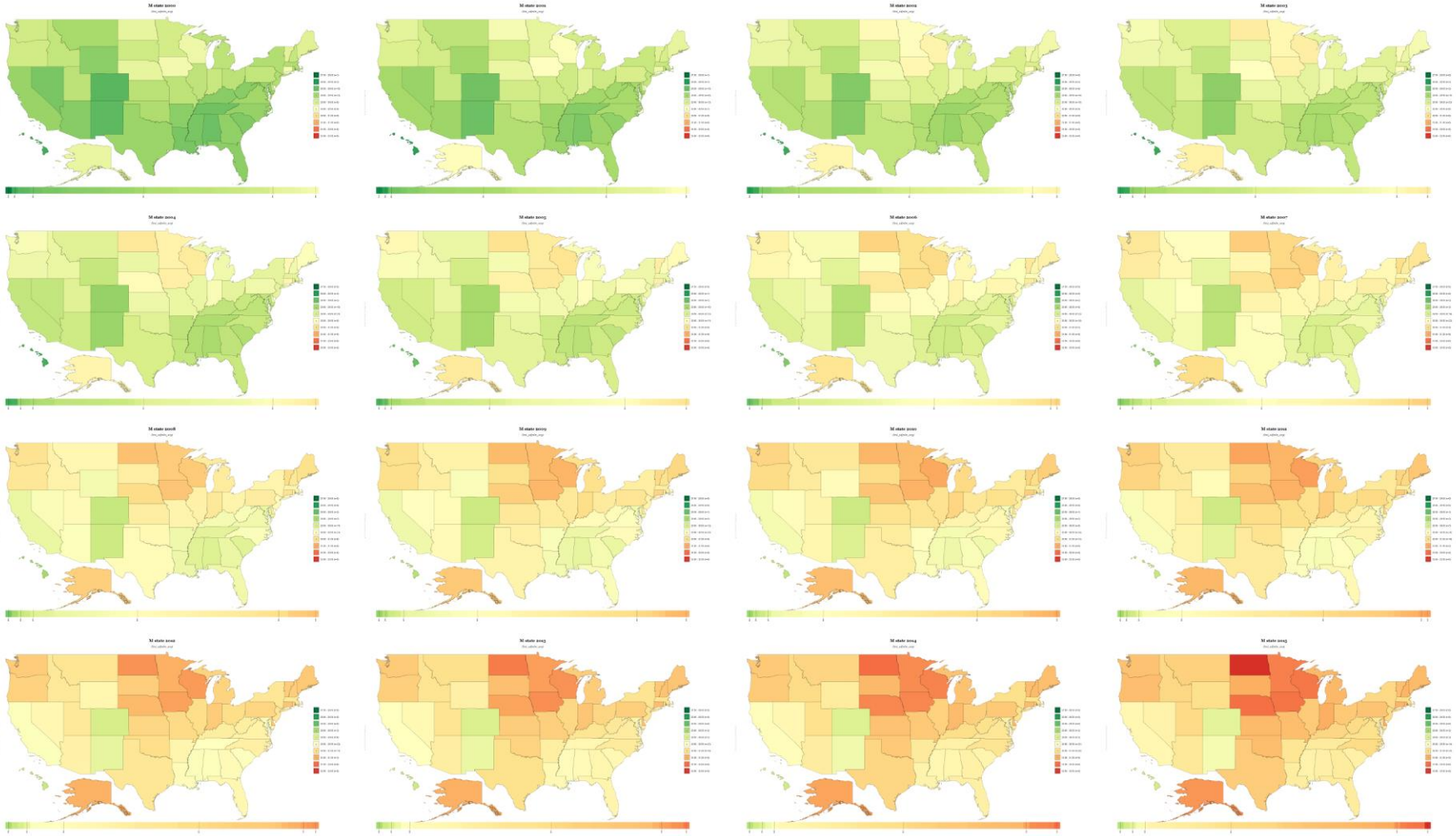
VA Male BMI Age Adjusted [MLM] – 2015



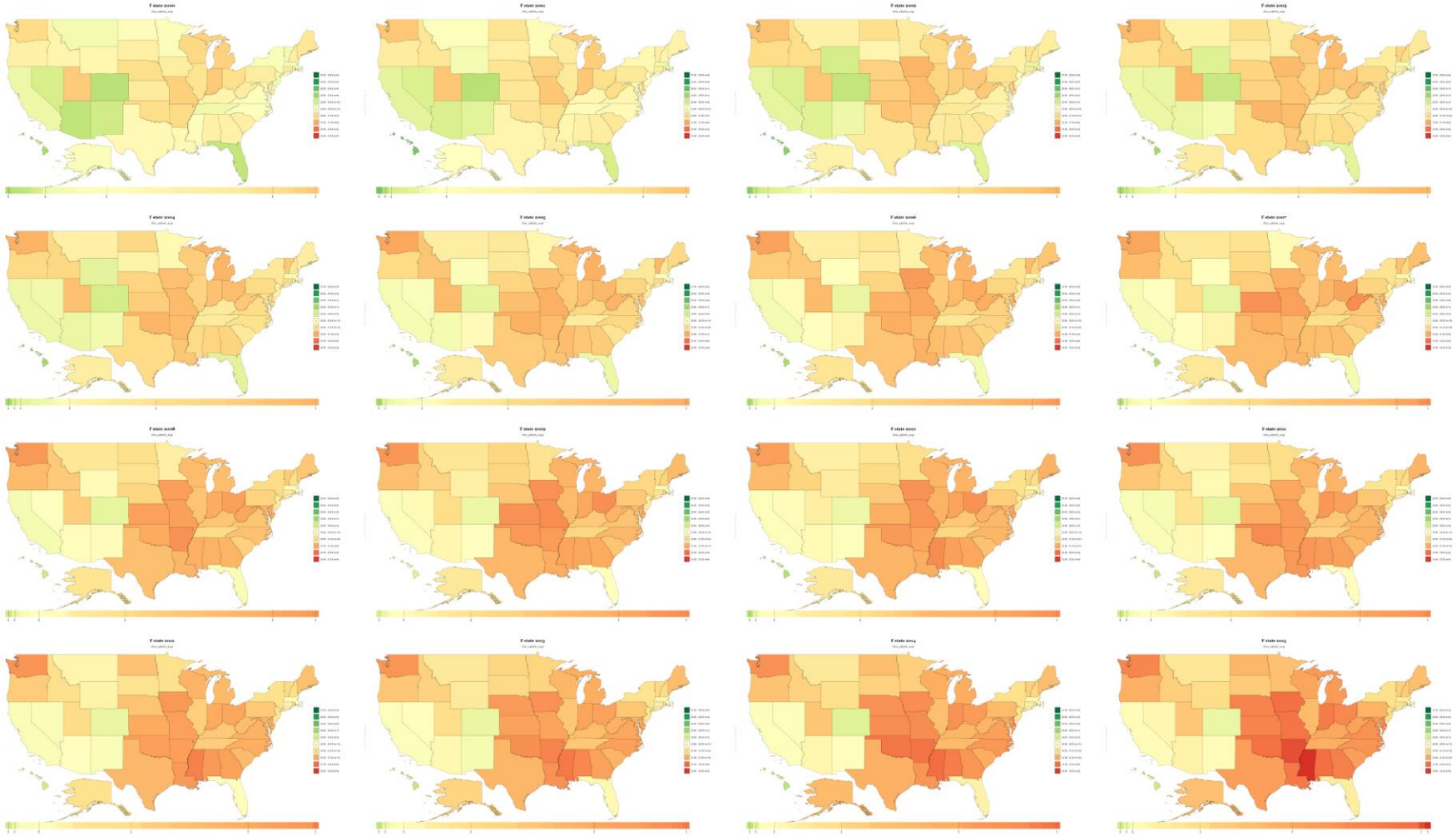
County BMI [MLM] Males 2000-2015



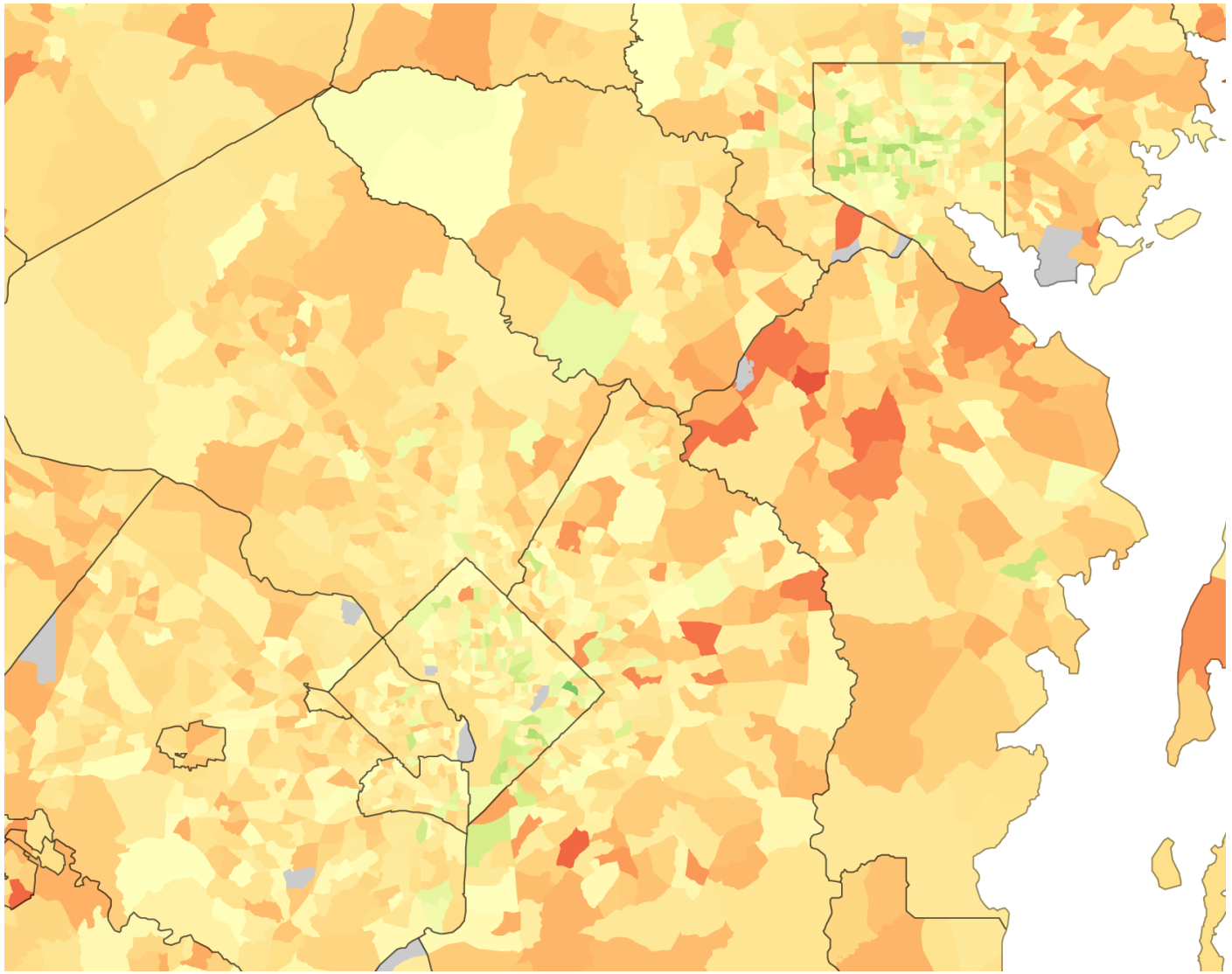
State BMI [MLM] Males 2000-2015



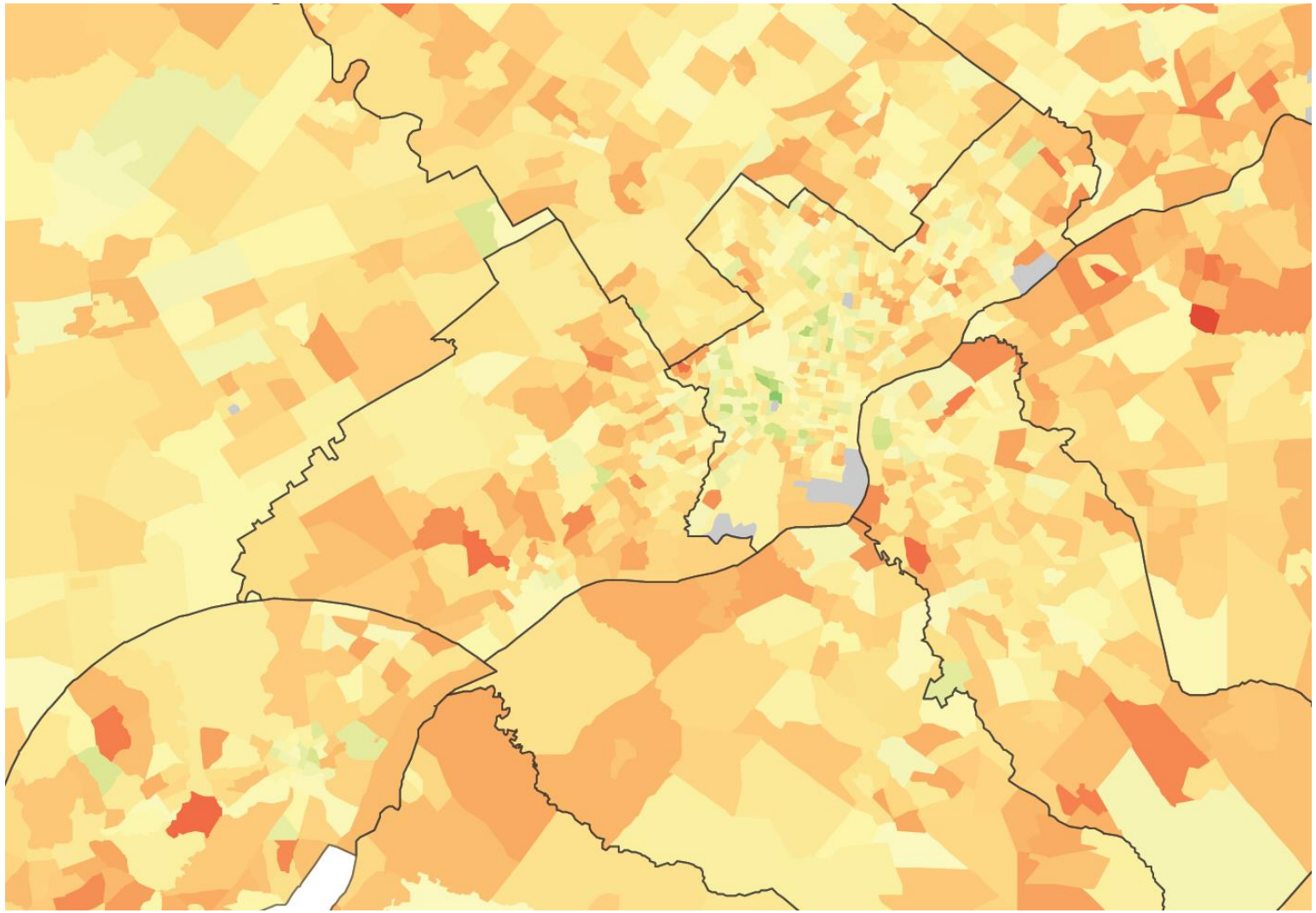
State BMI [MLM] Females 2000-2015



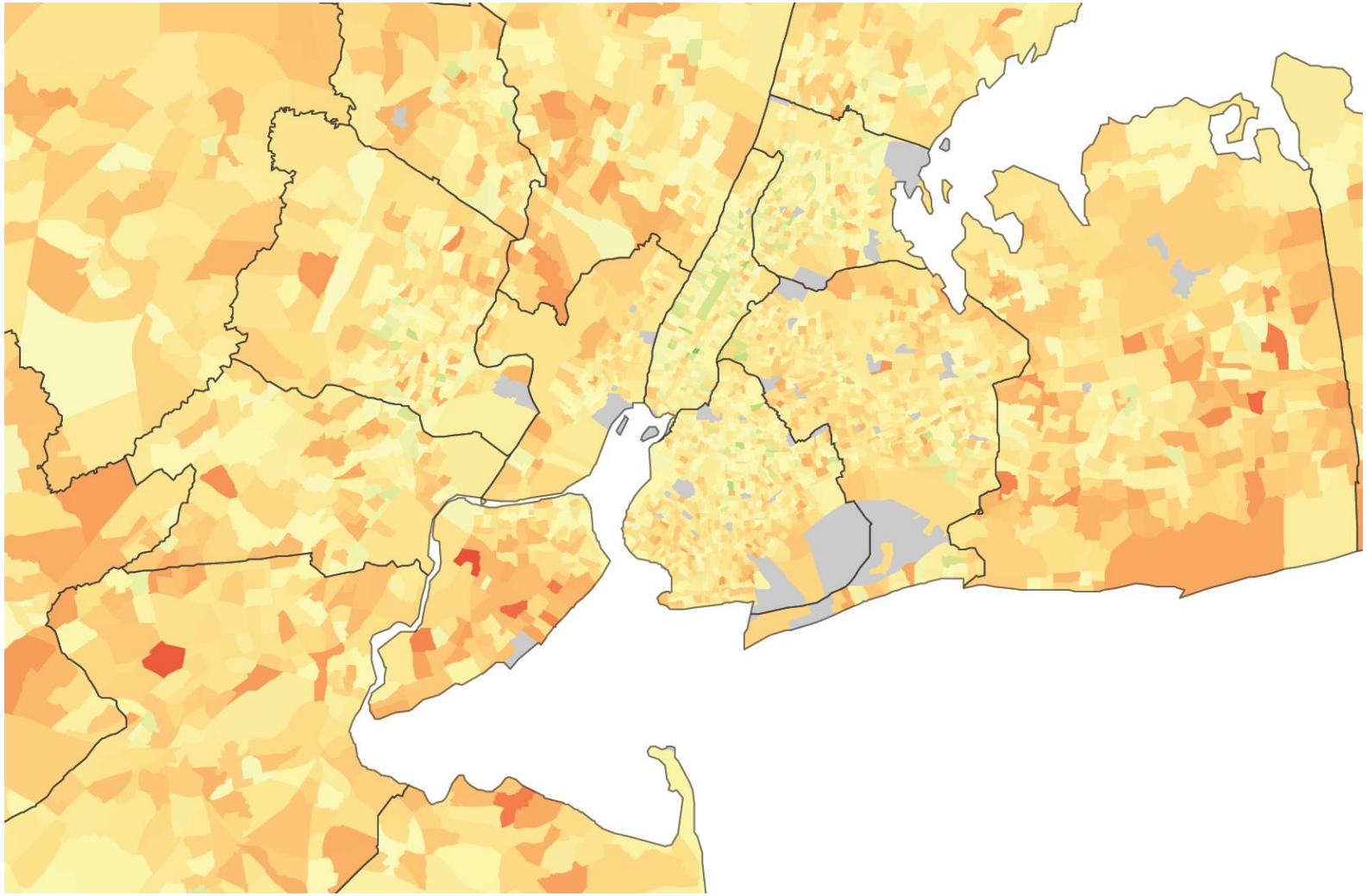
Tract BMI [MLM] Male (DC/Baltimore)



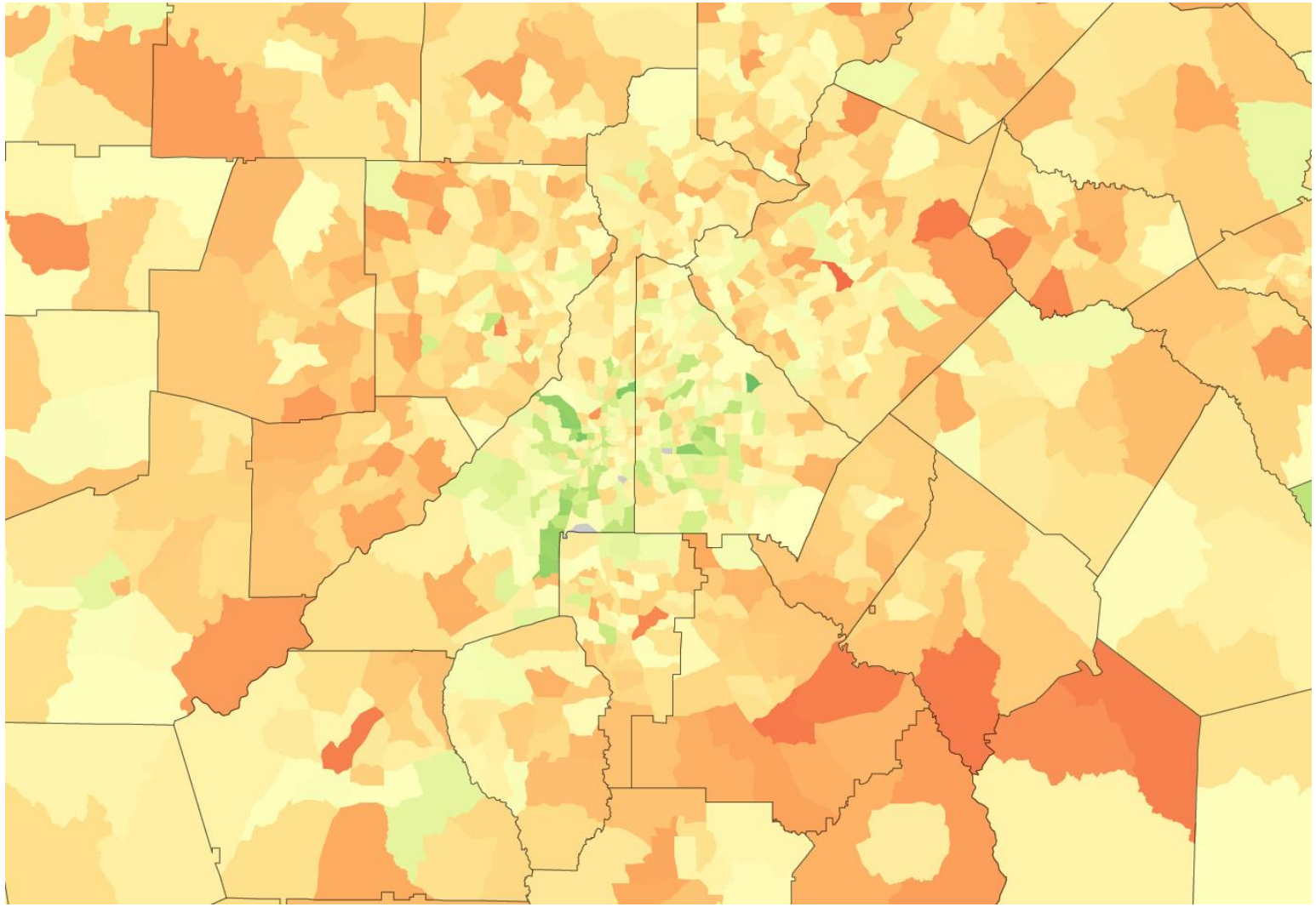
Tract BMI [MLM] Male (Philadelphia)



Tract BMI [MLM] Male (New York City)

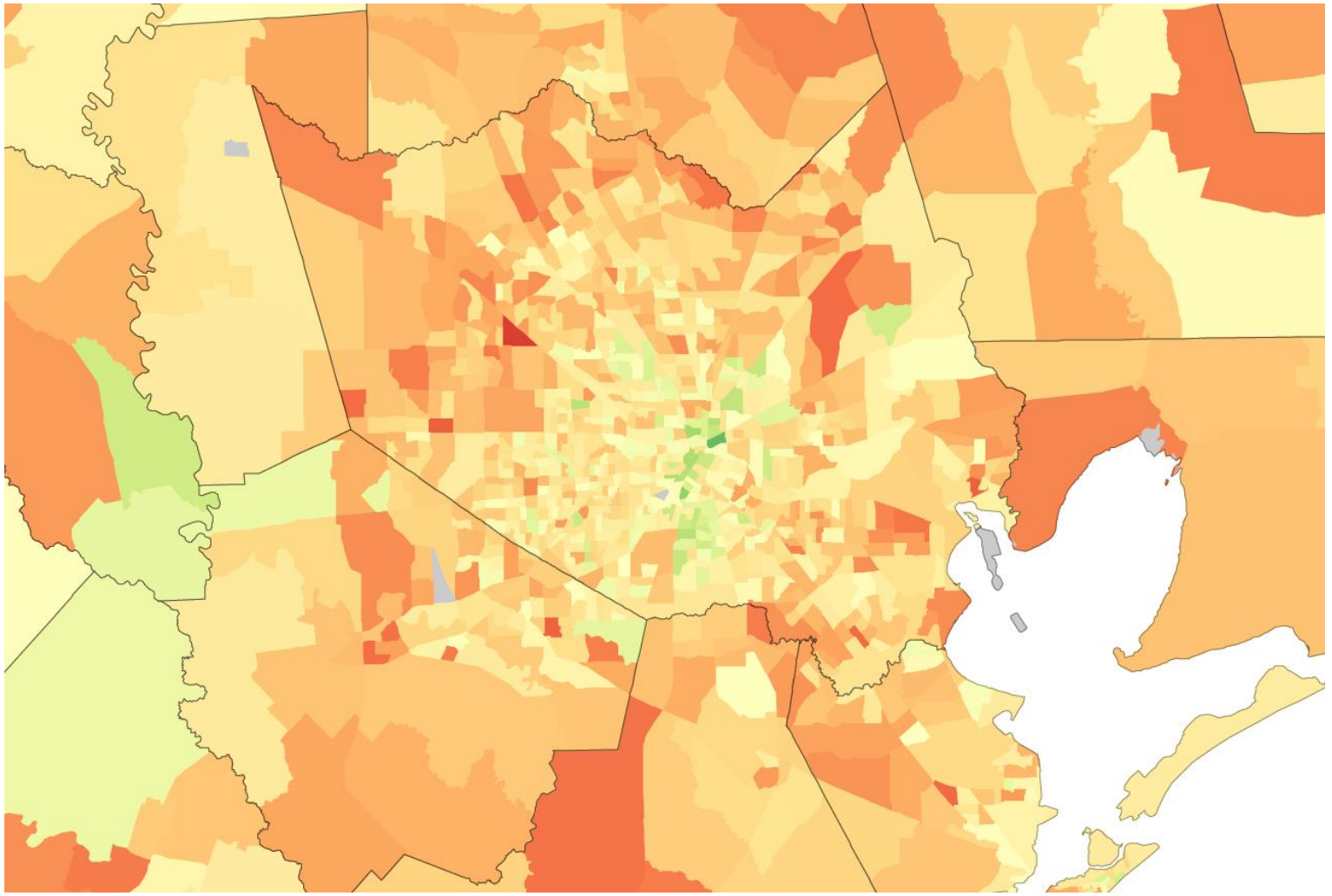


Tract BMI [MLM] Male (Atlanta)

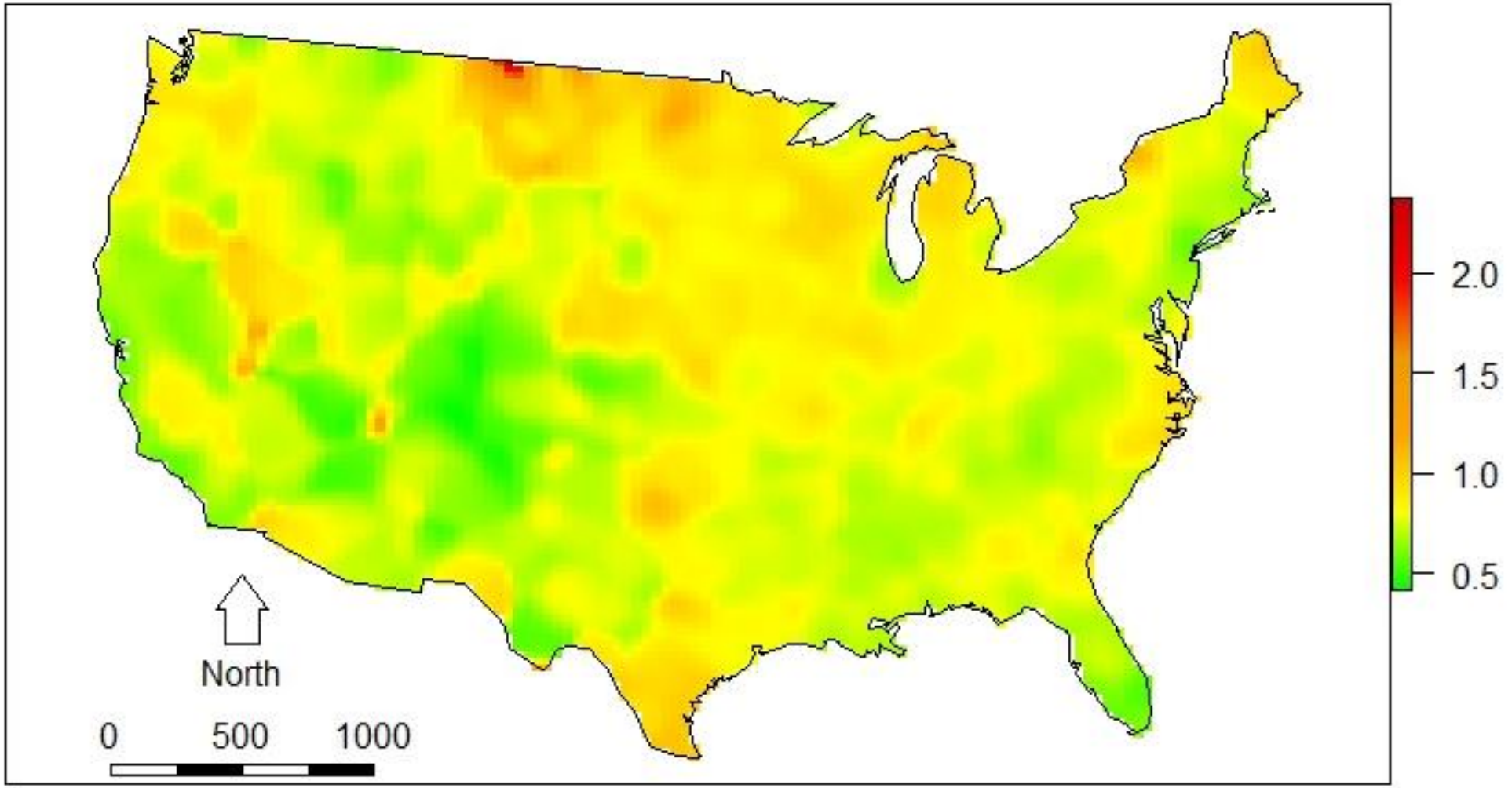


Tract BMI [MLM] Male (Chicago)

Tract BMI [MLM] Male (Houston)



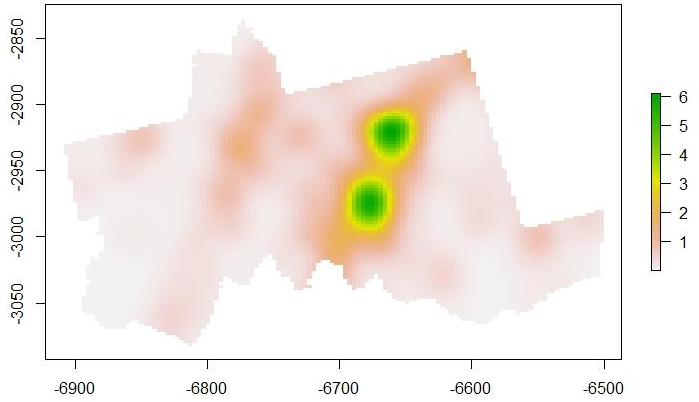
Spatial Intensity Male 2015

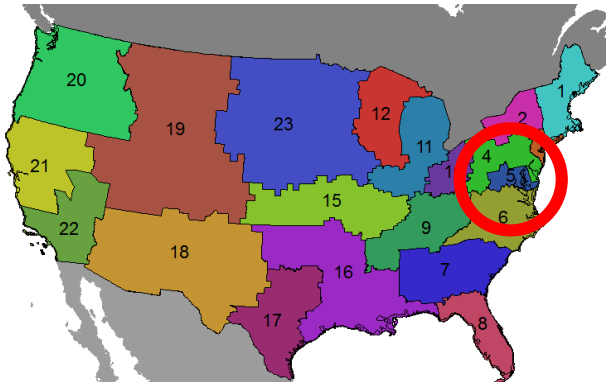


VISN #5: VA Capitol Healthcare Network

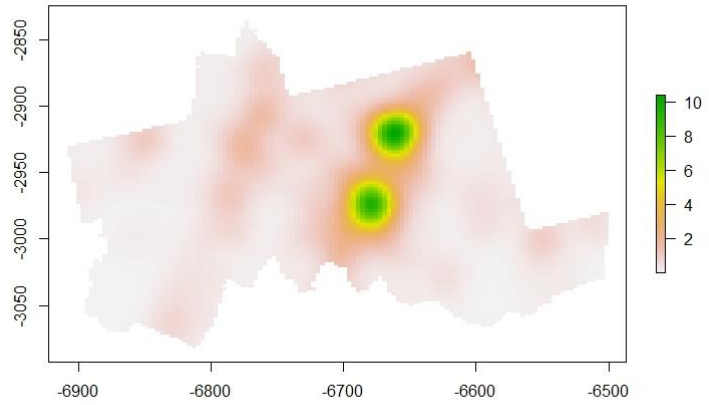
(Maryland, Virginia, Washington, D.C., West Virginia)

Spatial intensity of cases

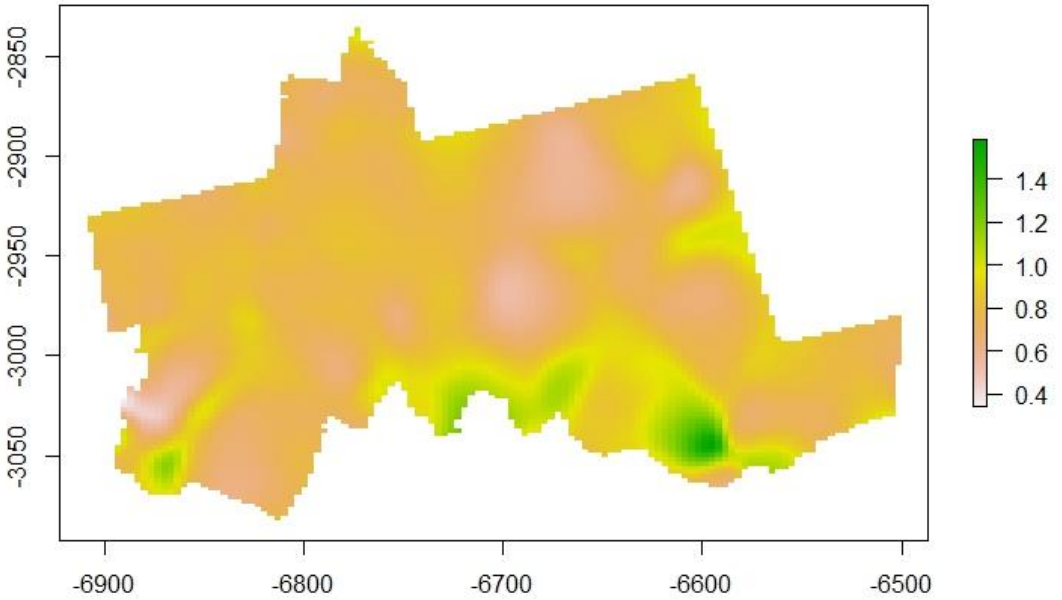




Spatial intensity of non-cases



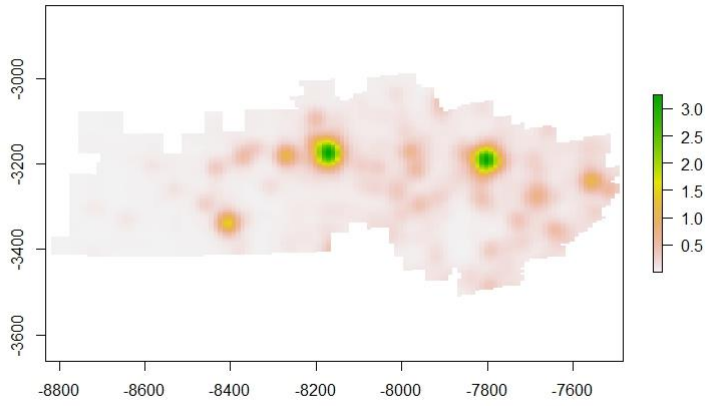
Spatial odds for obesity

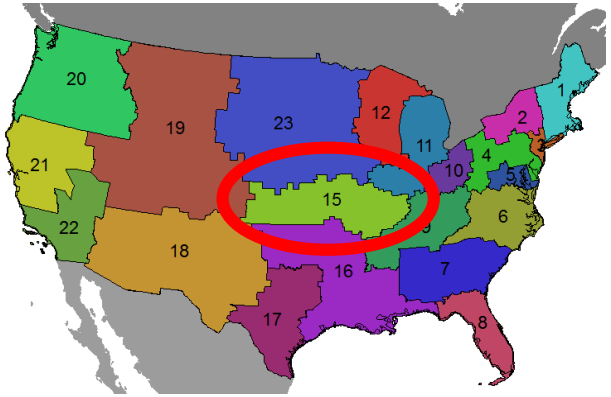


Bandwidth: 10 km

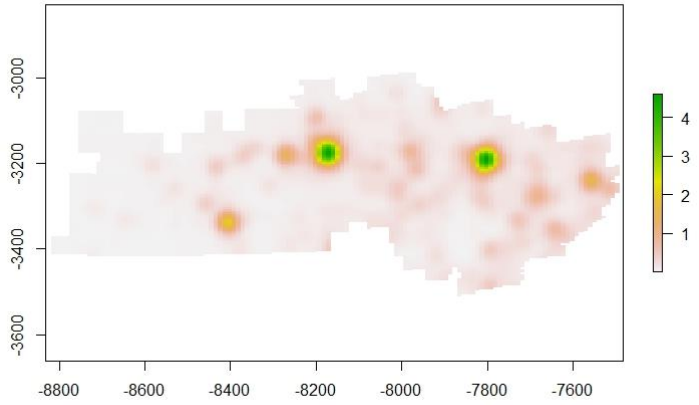
VISN #15: VA Heartland Network
(Illinois, Kansas, Missouri)

Spatial intensity of cases

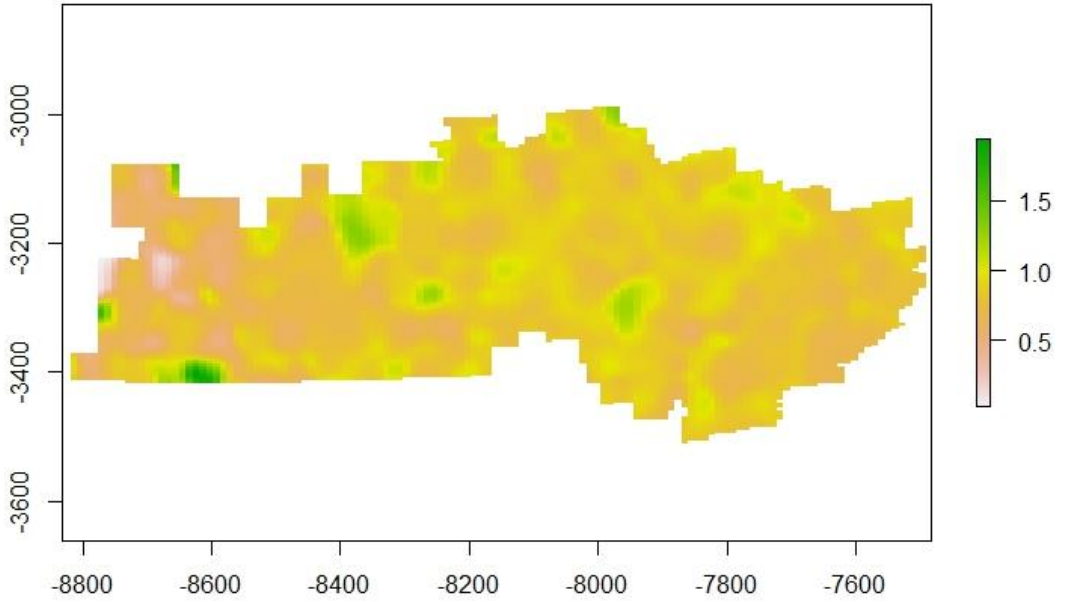




Spatial intensity of non-cases



Spatial odds for obesity

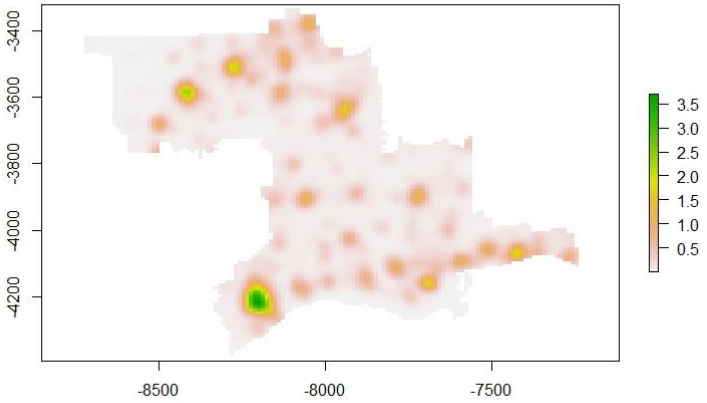


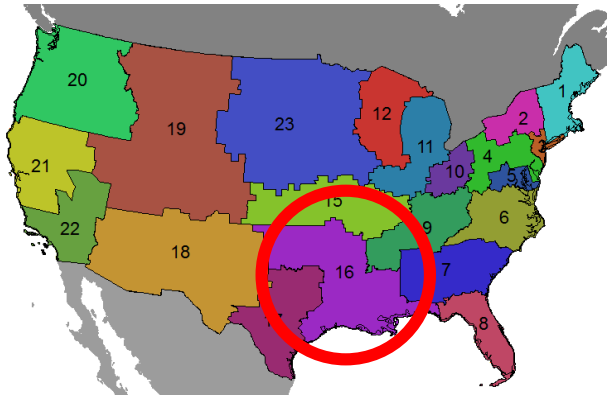
Bandwidth: 15 km

VISN #16: South Central VA Healthcare Network

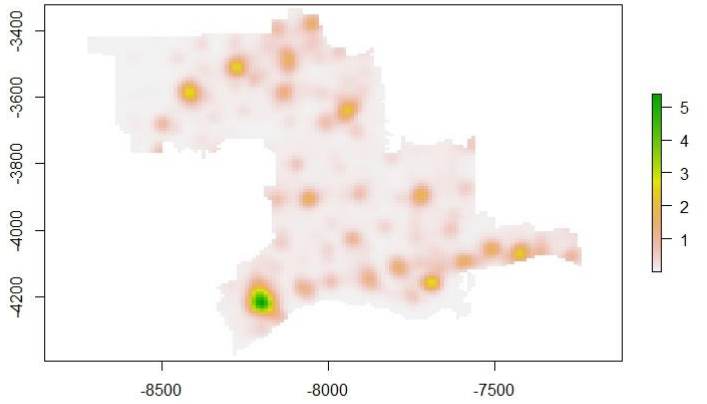
(Arkansas, Louisiana, Mississippi, Oklahoma, Texas)

Spatial intensity of cases

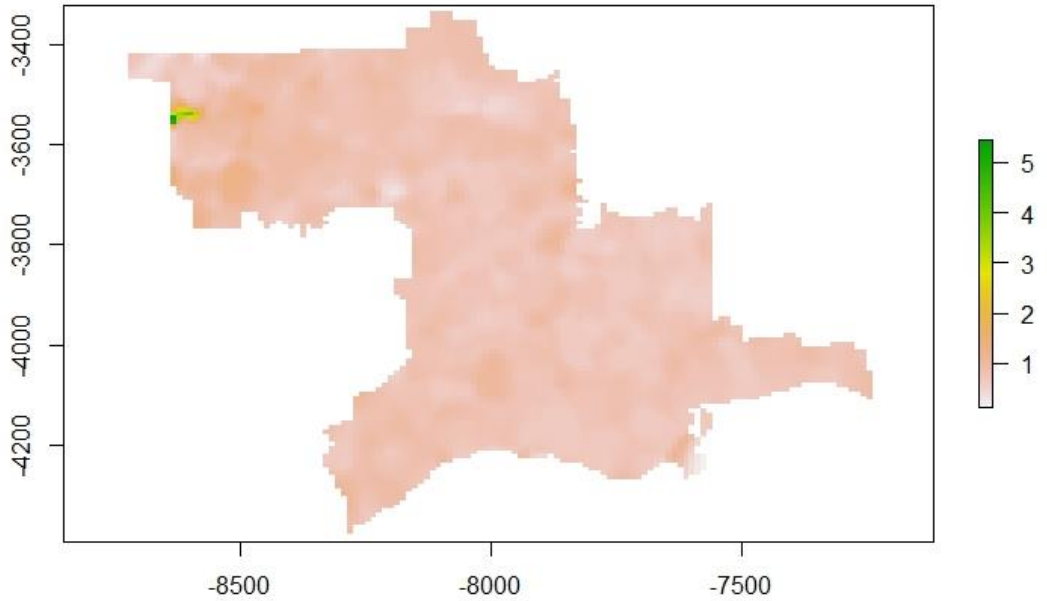




Spatial intensity of non-cases



Spatial odds for obesity

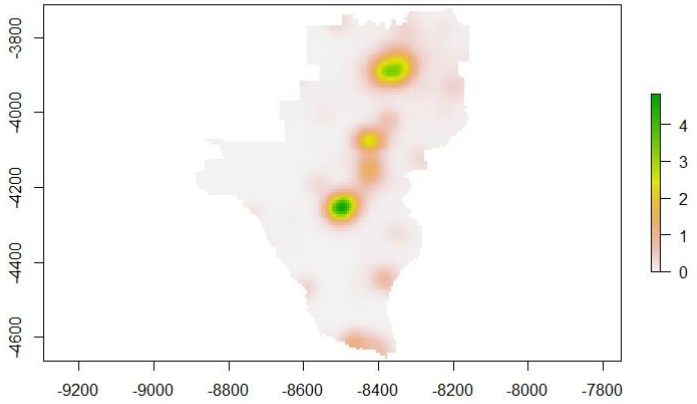


Bandwidth: 15 km

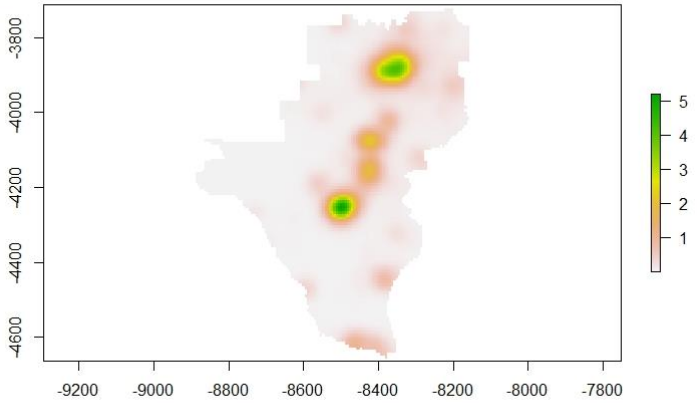
VISN #17: VA Heart of Texas Healthcare System

(Texas)

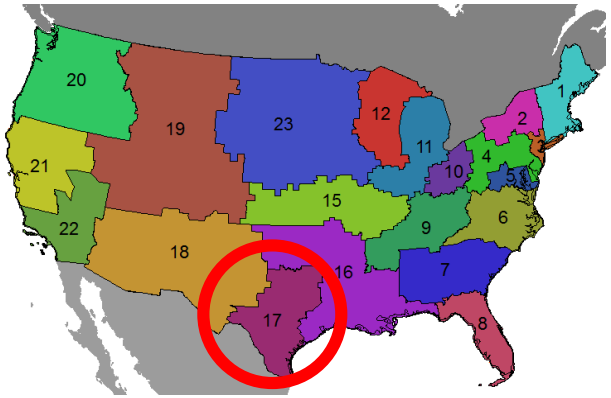
Spatial intensity of cases



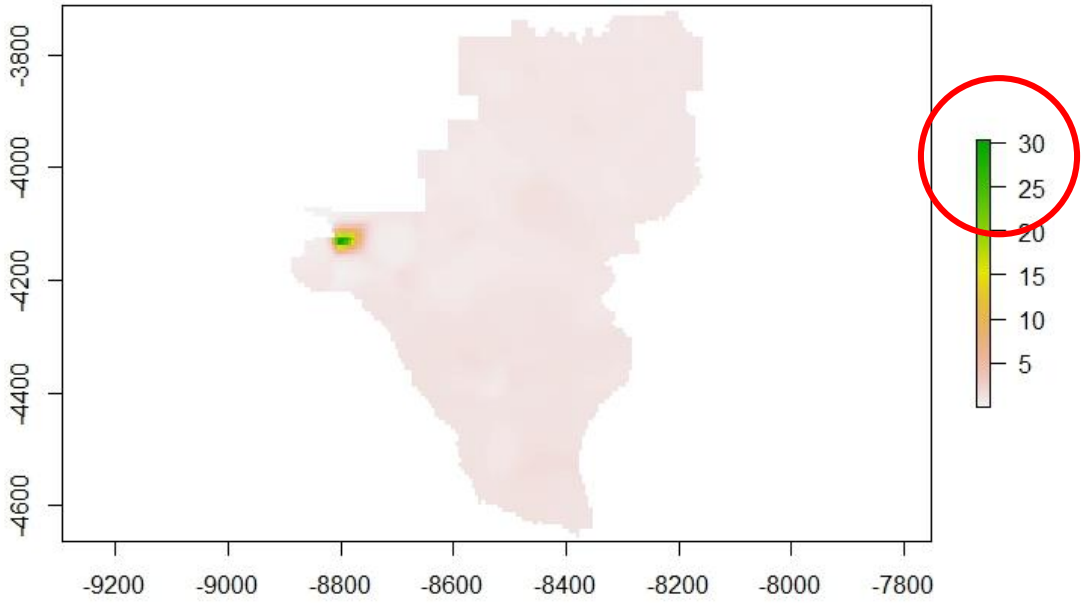
Spatial intensity of non-cases



Bandwidth: 20 km



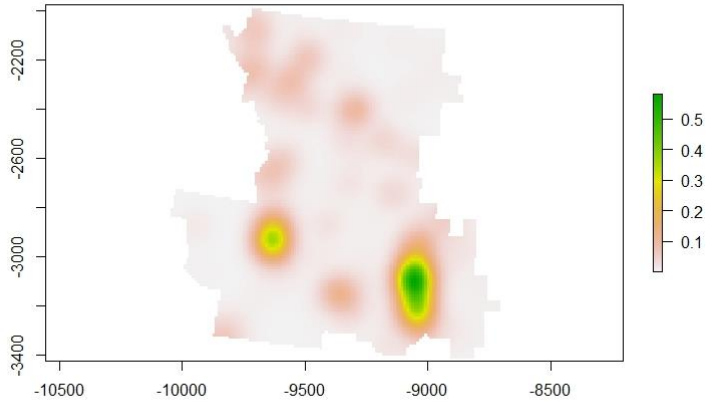
Spatial odds for obesity

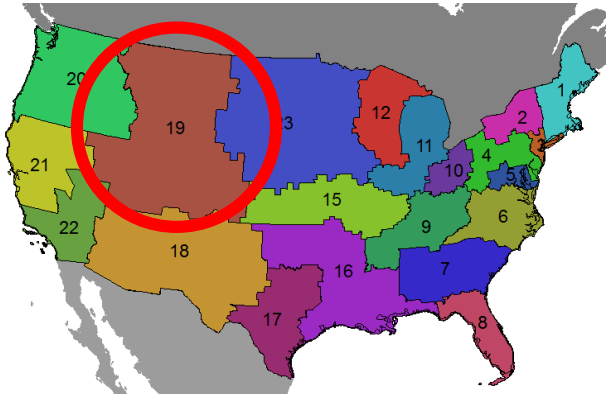


VISN #19: VA Rocky Mountain Network

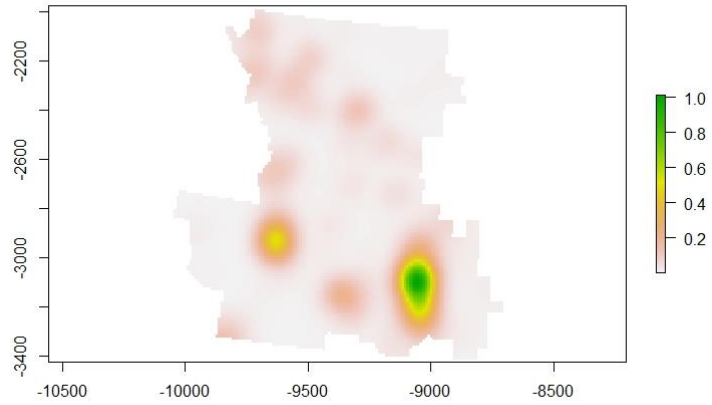
(Colorado, Montana, Utah, Wyoming)

Spatial intensity of cases

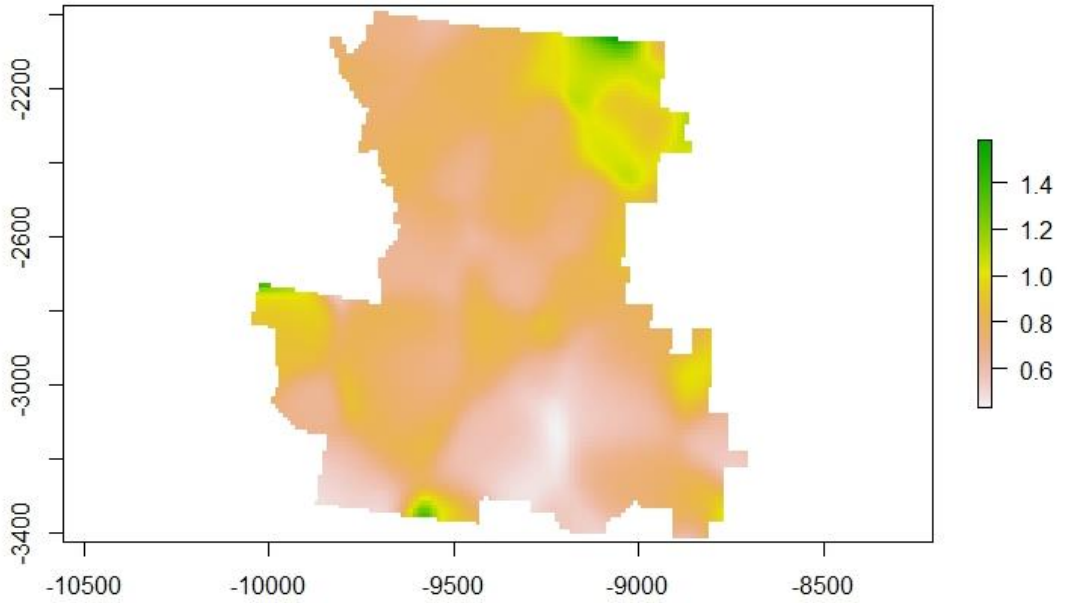




Spatial intensity of non-cases

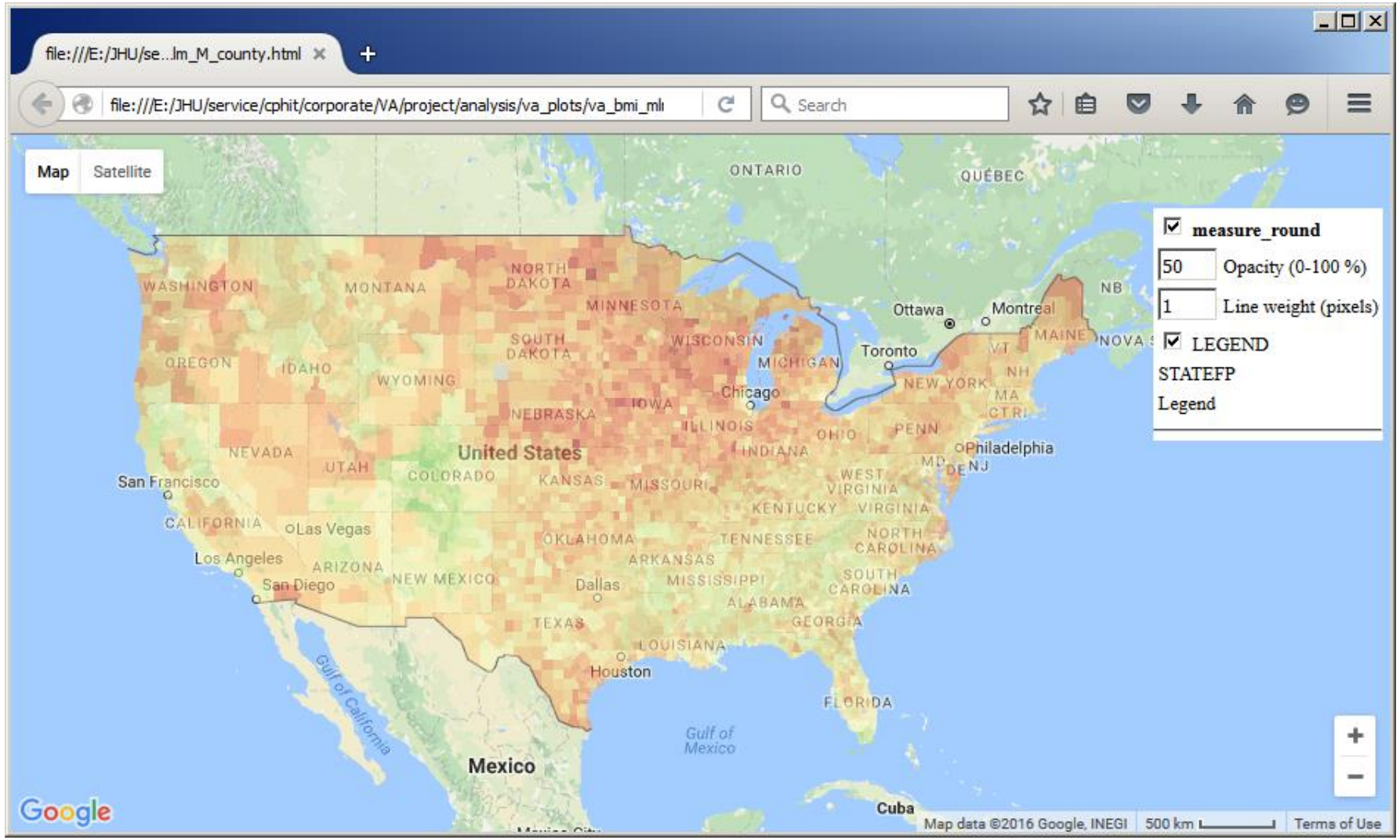


Spatial odds for obesity

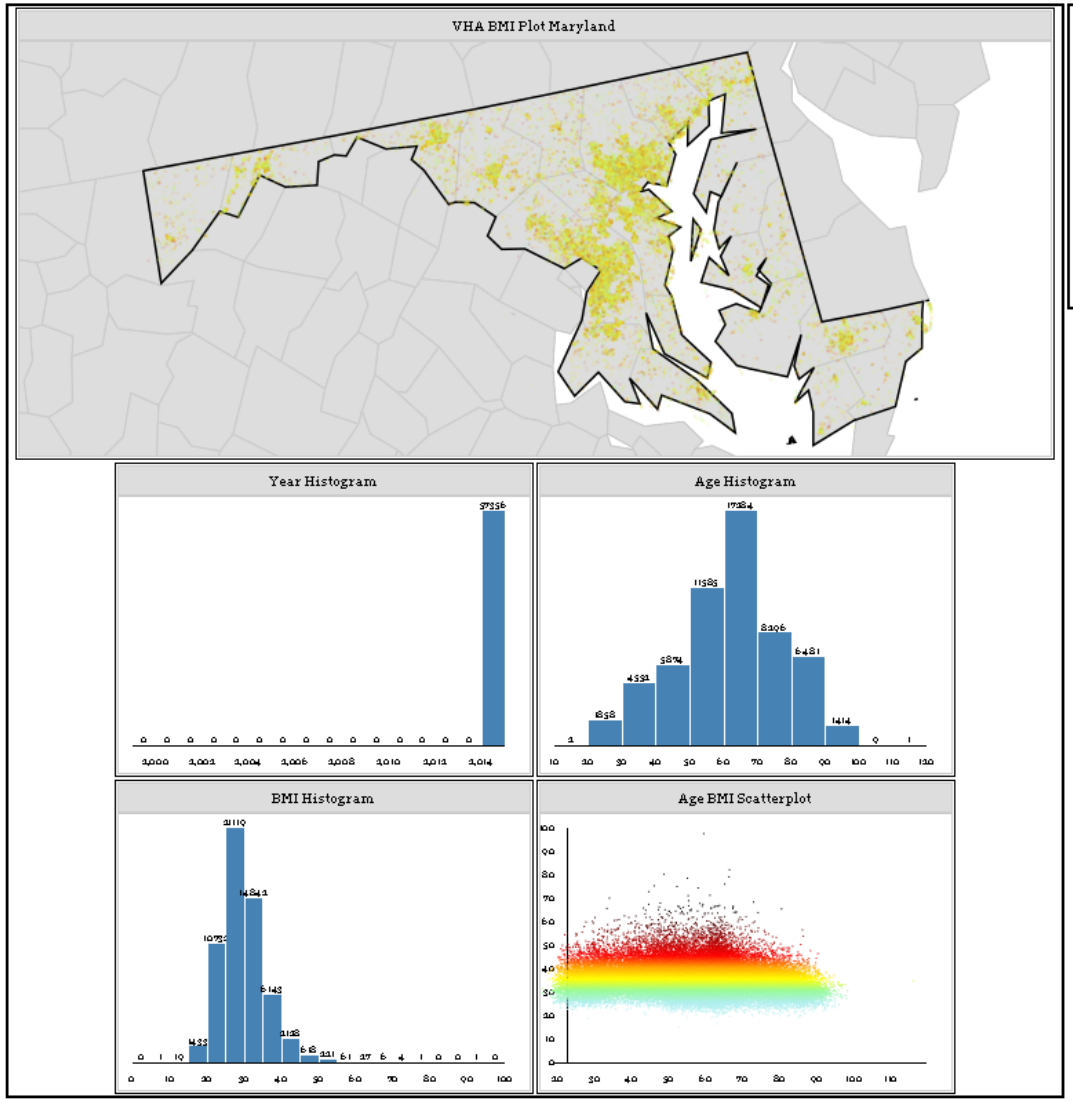


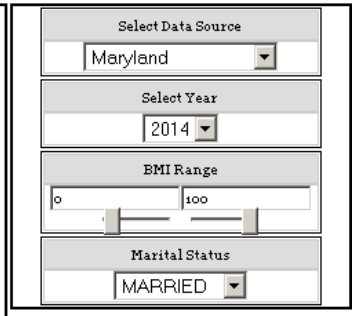
Bandwidth: 50 km

Interactive Visualization



Visual Analytics





Name	Owner
VHA Corporate Data Warehouse	VHA
American Community Survey	Census
Census 2010	Census
National Health and Nutrition Examination Survey	CDC
Food Access Research Atlas + Others	USDA
National Vital Statisitcs Report	CDC
Reference USA	RefUSA
Open Street Map	OpenMap
Moderate Resolution Imaging Spectroradiometer	NASA
Consumer Expenditure Survey	BLS
Uniform Crime Reporting Statistics (FBI)	FBI
Maryland Food Systems	MD
USDA Detailed Maps Baltimore	USDA
ArcGIS Internal Datasets	ESRI
Satellite data	Google

Interactive Web-based Real-time Geo-Temporal Exploration of Obesity Data
(Showing averages of 2014 for MD)

Factors Associated with Obesity (Mid-Atlantic VISN)

Variable	Reference / Type	OR	p value	Increases Obesity
Age	Continuous	0.97	<0.001	lower age
Race (white)	Categorical (ref: non-white)	1.02	0.74	race = white
Income (< $25k)	Categorical (ref: > $25k)	0.88	<0.05	income > $25k
Marriage (not married)	Categorical (ref: married)	0.77	<0.001	married
Service years	Continuous	1.02	<0.001	more service year
SES Q2	Categorical (ref: Q1 lowest)	1.21	<0.05	higher SES quartile
SES Q3	Categorical (ref: Q1 lowest)	1.34	<0.001	higher SES quartile
SES Q4	Categorical (ref: Q1 lowest)	1.16	0.06	higher SES quartile
Gagne (> 0)	Categorical (ref: <= 0)	0.94	0.16	Gagne < 0
Urban/Rural (rural)	Categorical (ref: urban)	0.93	0.27	being urban
Road Density	Continuous	0.91	0.44	lower road density
Low Food Access	Continuous	1.06	0.19	higher low-food-access

Mid-Atlantic VISN – Multivariate GEE Analysis

high income and not married and high SES quartile remained increased odds of obesity,
but Gagne scores were no longer associated with obesity

Discussion

Poll Question #3

- In your opinion, what was the most time-consuming task in this population-level analysis research? (select one)
  - Cleaning weight and height data
  - Attaching geo-data to underlying populations
  - Developing models to predict obesity trajectory
  - Understanding the factors associated with obesity

Challenges and Opportunities in Population Health Analytic

- **Data sources/types**
  - How to compare data types and their added value?
  - What are the limits of each data type? What are we missing?
  - What can be used from unstructured data?
- **Data quality**
  - How much juice is left in this data type (e.g., claims)?
  - Do objective measures have data quality issues (e.g., BMI)?
  - How can we measure the quality of subjective data?
- **Denominator**
  - Are we excluding noise or signal?
  - Is this a too big of a cut or too narrow – sample size issues?
  - Patient attribution issues…

Challenges and Opportunities of Using EHR Data

- **Challenges**
  - EHRs are transactional and extracting temporal data for population health research is complex
  - EHRs have limited data quality and require extensive preparation before use

- **Opportunities**
  - EHRs can be a viable/reliable source of data for population health research (e.g., objective measure of W/H instead of self-reported)
  - EHRs provide almost real-time data
  - EHRs save efforts in collecting population-wide data (e.g., replacing surveys); however, population denominator adjustments might be challenging at the same time

Potential VHA Collaborations

❑ Going Beyond CDW (new data sources/types)

- Assess the role of social determinants of health (both personal level and geo-derived) in VHA's population health research (e.g., patient phenotyping; controlling for or measuring the effect mediators and moderators; studying disparities)
- Evaluate the value of additional clinical data sources in VHA's population health research (e.g., *insurance claims through the Choice/MISSION Act; anticipated DoD Cerner data*)

❑ Assessing New Methods and Informatics Research

- Measuring the effect of data quality on population health research outcomes and findings
- Assessing the added-value of new methods (in relation to new data sources) in identifying and/or predicting certain population health outcomes (e.g., *temporal trend analysis; geo-analysis; NLP; machine learning methods*)
- Advancing the science of risk stratification within VHA's specific sub-populations (e.g., detecting specific patterns of high utilization within the mental health sub-populations)

Acknowledgement

- **VHA OABI / CSDE**
  - Dr. Stephan Fihn
  - Dr. Tamara Box
  - PACT PIs & team members
- **JHU CPHIT**
  - Dr. Jonathan Weiner
  - Dr. Frank Curriero
  - Dr. Elham Hatef
  - Dr. Kim Gudzune
  - Dr. Fardad Gharghabi
  - Kelly Searle
  - Elyse Lasser
  - Thomas Richards

(please contact kharazi@jhu.edu for the full list)

Poll Question #4

- In your opinion, what should be the next step for VHA's population health research? (select all that apply)
  - Developing the technical infrastructure (considering the EHR migration)
  - Assessing the value of various data sources (e.g., geo-data, claims data)
  - Identifying and targeting disparities (e.g., mental health, females)
  - Funding future studies to exemplify population-level analysis

Thank you!

Q & A

Hadi Kharrazi *MD PhD*
kharrazi@jhu.edu
www.jhsph.edu/cphit