County CCPD Index via Principal Component Analysis

Benjamin White, MPH

Renee Calanan, Ph.D

Elisabeth Meyer, MPH

Devon Williford, MPH

March 2017





Previous Strategy: "Rank the Rankings"





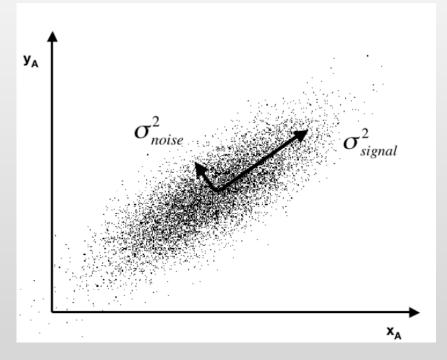
Principal Components Analysis

A type of factor analysis

A data reduction process

Filters out data "noise" and reduces data down to components or "factors" of combined variables.

Factors are the underlying structure of the data and explain the most variance in a dataset.





Principal Components Example: CO Craft Beer











Examples of PCA-Based Indices

Journal of Urban Health: Bulletin of the New York Academy of Medicine, Vol. 83, No. 6 doi:10.1007/s11524-006-9094-x

© 2006 The New York Academy of Medicine

The Development of a Standardized Neighborhood Deprivation Index

Lynne C. Messer, Barbara A. Laraia, Jay S. Kaufman, Janet Eyster, Claudia Holzman, Jennifer Culhane, Irma Elo, Jessica G. Burke, and Patricia O'Campo RESEARCH

Krefis et al. Malaria Journal 2010, 9:201

MALARIA JOURNAL

Open Access

Principal component analysis of socioeconomic factors and their association with malaria in children from the Ashanti Region, Ghana

Anne Caroline Krefis^{1*}, Norbert Georg Schwarz¹, Bernard Nkrumah³, Samuel Acquah³, Wibke Loag¹, Nimako Sarpong³, Yaw Adu-Sarkodie⁴, Ulrich Ranft², Jürgen May¹

"Measuring early childhood health: a composite index comparing Colombian departments"

Ana María Osório, Catalina Bolancé and Manuela Alcañiz

© The Author 2006, Published by Oxford University Press in association with The London School of Hygiene and Tropical Medicine. All rights reserved. doi:10.1093/heapoliczi029 Advance Access publication 9 October 2006

How to do (or not to do)...

Constructing socio-economic status indices: how to use principal components analysis

SEEMA VYAS AND LILANI KUMARANAYAKE HIVTools Research Group, Health Policy Unit, Department of Public Health and Policy, London School of Hygiene and Tropical Medicine, London, UK



Index Steps 🖪

- Determine which variables to include
- Clean data, fill in missing data via statistical methods
- Standardize all variables (counts, incidence, proportions)
- Look for correlations among variables to better understand data
- Run Principal Components Analysis in **SSAS**
- Construct weights for components (factors) based on explained variance and apply them
- Create the Index and rank index scores 1-64 (lower # = better)



Index Data Sources









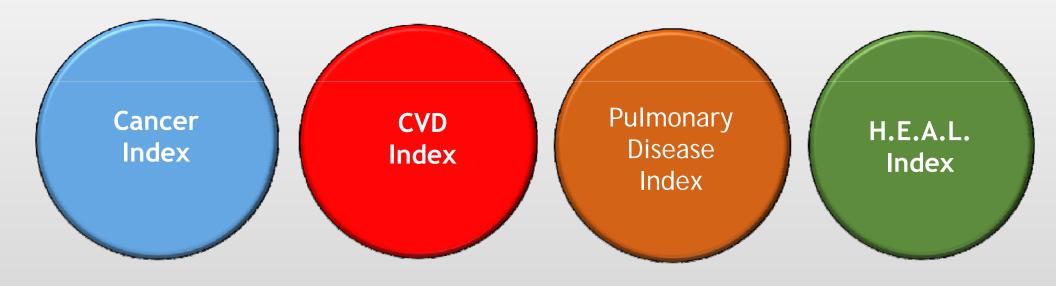








Indices Created





Cancer Index: Rankings



All cancers combined, crude incidence rate per 100,000 persons (2013-2015, Colorado Central Cancer Registry (CCCR))

All cancers combined, crude prevalence (%) (2013-2015, CCCR)

All cancers combined, crude mortality rate per 100,000 persons (2013-2015, vital statistics)

Cervical cancer, crude incidence rate per 100,000 persons (2013-2015, CCCR)

Melanoma, crude incidence rate per 100,000 persons (2013-2015, CCCR)

Colorectal cancer, crude incidence rate per 100,000 persons (2013-2015, CCCR)

Lung cancer, crude incidence rate per 100,000 persons (2013-2015, CCCR)

Colorectal cancer, % late stage diagnosis (2013-2015, CCCR)

Lung cancer, % late stage diagnosis (2013-2015, CCCR)

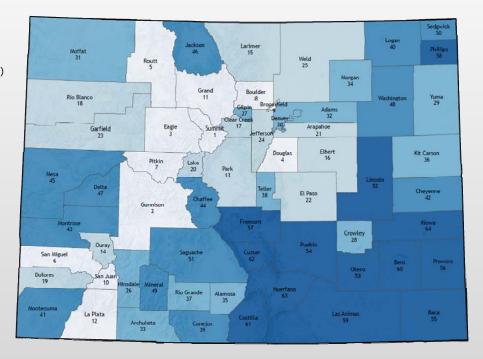
Percent of Population Minority Race or Hispanic Latino Ethnicity (U.S. Census ACS 5 yr., 2011-2015)

Percent of the Population Age 25+ Without a H.S Diploma (U.S. Census ACS 5 yr., 2011-2015)

Percent of the Total Noninstitutionalized Civilian Population that is Uninsured (U.S. Census ACS 5 yr., 2011-2015)

Percent of population on Medicaid (U.S. Census ACS 5 yr., 2011-2015)

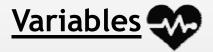
Percent of population below annual Income Self Sufficiency Standard needed for a family with 2 adults, One preschooler and one school-age child (2015 Colorado Self Sufficiency Report; U.S. Census ACS 5 yr., 2011-2015)







Cardiovascular Disease Index: Rankings



Percent of adults who have been diagnosed with high blood pressure (2013-2015, BRFSS)

Percent of adults aged 20+ years who have been diagnosed with high cholesterol (2013 & 2015, BRFSS)

Percent of adults who have been diagnosed with angina/CHD, heart attack, or stroke (2013-2015, BRFSS)

Percent of adults who have been diagnosed with diabetes (2013-2015, BRFSS)

Diabetes hospital discharge rate per 100,000 (Colorado Hospital Association, 2013-2015)

Diabetes emergency department visits per 100,000 (Colorado Hospital Association, 2013-2015)

Diabetes premature (before age 75) mortality rate per 100,000 (2013-2015, vital statistics)

Major Cardiovascular Disease hospital discharge rate per 100,000 (Colorado Hospital Association, 2013-2015)

Major Cardiovascular Disease emergency department visits per 100,000 (Colorado Hospital Association, 2013-2015)

Major cardiovascular disease premature (before age 75) mortality rate per 100,000 (2013-2015, vital statistics)

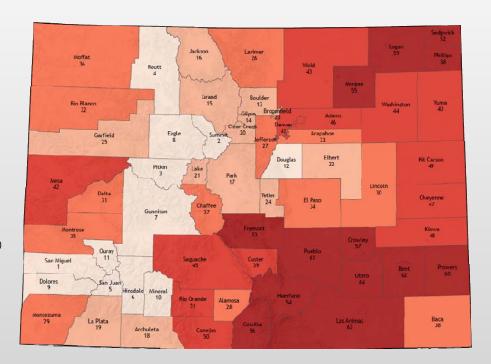
Percent of Population Minority Race or Hispanic Latino Ethnicity (U.S. Census ACS 5 yr., 2011-2015)

Percent of the Population Age 25+ Without a H.S Diploma (U.S. Census ACS 5 yr., 2011-2015)

Percent of the Total Noninstitutionalized Civilian Population that is Uninsured (U.S. Census ACS 5 yr., 2011-2015)

Percent of population on Medicaid (U.S. Census ACS 5 yr., 2011-2015)

Percent of population below annual Income Self Sufficiency Standard needed for a family with 2 adults, One preschooler and one school-age child (2015 Colorado Self Sufficiency Report; U.S. Census ACS 5 yr., 2011-2015)







Pulmonary Disease Index: Rankings

Variables &

Percent of children who currently have asthma (2013-2015, Child Health Survey)

Percent of high school students who have ever been diagnosed with asthma (2015, Healthy Kids Colorado Survey)

Percent of adults who currently have asthma (2013-2015, BRFSS)

Percent of adults who have been diagnosed with COPD (2013-2015, BRFSS)

Asthma Hospital discharge rate per 100,000 (Colorado Hospital Association, 2013-2015)

Asthma Emergency Department Visits per 100,000 (Colorado Hospital Association, 2013-2015)

Chronic Lower Respiratory Disease Hospital discharge rate per 100,000 (Colorado Hospital Association, 2013-2015)

Chronic Lower Respiratory Disease Emergency Dept. Visits per 100,000 (Colorado Hospital Association, 2013-2015)

Chronic Lower Respiratory Disease (exclusive of asthma) mortality rate per 100,000 (2013-2015, vital statistics)

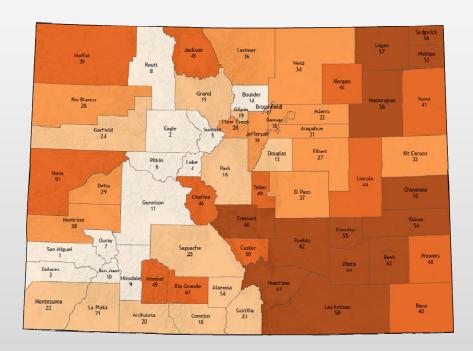
Percent of Population Minority Race or Hispanic Latino Ethnicity (U.S. Census ACS 5 yr., 2011-2015)

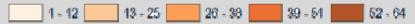
Percent of the Population Age 25+ Without a H.S Diploma (U.S. Census ACS 5 yr., 2011-2015)

Percent of the Total Noninstitutionalized Civilian Population that is Uninsured (U.S. Census ACS 5 yr., 2011-2015)

Percent of population on Medicaid (U.S. Census ACS 5 yr., 2011-2015)

Percent of population below annual Income Self Sufficiency Standard needed for a family with 2 adults, One preschooler and one school-age child (2015 Colorado Self Sufficiency Report; U.S. Census ACS 5 yr., 2011-2015)







H.E.A.L. Index: Rankings



Percent of adults who are obese (2013-2015, BRFSS)

Percent of high school students who are obese (2015, Healthy Kids Colorado Survey (HKCS))

Percent of children aged 5-14 who are obese (2013-2015, Child Health Survey)

Percent of adults who are overweight (2013-2015, BRFSS)

Percent of high school students who are overweight (2015, Healthy Kids Colorado Survey (HKCS))

Percent of children aged 5-14 who are overweight (2013-2015, Child Health Survey)

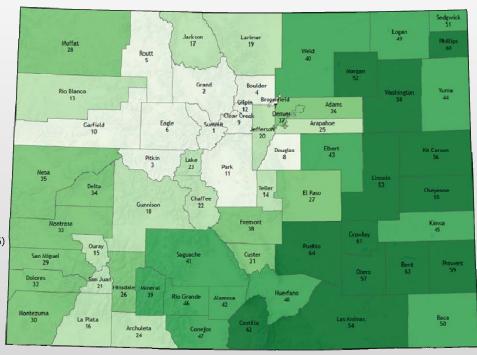
Percent of Population Minority Race or Hispanic Latino Ethnicity (U.S. Census ACS 5 yr., 2011-2015)

Percent of the Population Age 25+ Without a H.S Diploma (U.S. Census ACS 5 yr., 2011-2015)

Percent of the Total Noninstitutionalized Civilian Population that is Uninsured (U.S. Census ACS 5 yr., 2011-2015)

Percent of population on Medicaid (U.S. Census ACS 5 yr., 2011-2015)

Percent of population below annual Income Self Sufficiency Standard needed for a family with 2 adults, One preschooler and one school-age child (2015 Colorado Self Sufficiency Report; U.S. Census ACS 5 yr., 2011-2015)





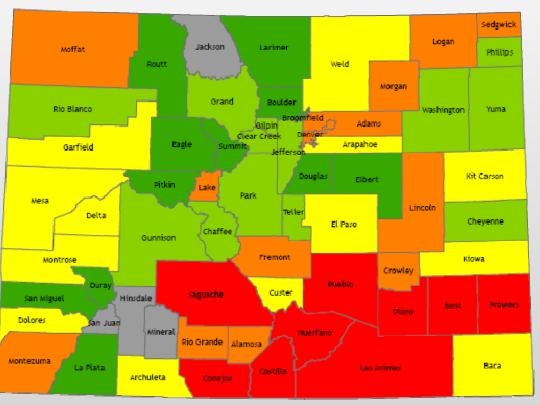


Comparisons: County Health Rankings (2015)

Health Factors		,	,
Health Behaviors	Clinical Care	Social & Economic Factors	Physical Environment
Adult smoking Adult obesity Food environment index Physical inactivity Access to exercise opportunities Excessive drinking Alcohol-impaired driving deaths Sexually transmitted infections Teen births	Uninsured Primary care physicians Dentists Mental health providers Preventable hospital stays Diabetes monitoring Mammography screening	Injury deaths	Air pollution - particulate matter Drinking water violations Severe housing problems Driving alone to work Long commute - driving alone



Comparison: County Health Rankings, Health Factors Rank (2015)





Obs/Vars Ratio: 2.1:1



COUNTY	Cancer Index Rank	CVD Index Rank	Pulmonary Index Rank	HEAL Index Rank
Adams	32	46	33	36
Alamosa	35	28	14	42
Arapahoe	21	33	31	25
Archuleta	33	18	20	24
Baca	55	38	43	50
Bent	60	62	63	63
Boulder	8	13	12	4
Broomfield	9	23	17	7
Chaffee	44	37	46	22
Cheyenne	42	47	52	55
Clear Creek	17	20	28	9
Conejos	39	50	18	47
Costilla	61	56	23	62
Crowley	28	57	55	61
Custer	62	39	50	31
Delta	47	31	29	34
Denver	30	41	35	37



COUNTY	Cancer Index Rank	CVD Index Rank	Pulmonary Index Rank	HEAL Index Rank	
Dolores	19	9	3	32	
Douglas	4	12	13	8	
Eagle	3	8	2	6	
El Paso	22	34	37	27	
Elbert	16	22	27	43	
Fremont	57	53	60	38	
Garfield	23	25	24	10	
Gilpin	27	14	19	12	
Grand	11	15	15	2	
Gunnison	2	7	11	18	
Hinsdale	26	6	9	26	
Huerfano	63	54	61	48	
Jackson	46	16	45	17	
Jefferson	24	27	38	20	
Kiowa	64	48	54	45	
Kit Carson	36	49	32	56	
La Plata	12	19	21	16	



COUNTY	Cancer Index Rank	CVD Index Rank	Pulmonary Index Rank	HEAL Index Rank
Lake	20	21	4	23
Larimer	15	26	36	19
Las Animas	59	63	59	54
Lincoln	52	30	44	53
Logan	40	59	57	49
Mesa	45	42	51	35
Mineral	49	10	49	39
Moffat	31	36	39	28
Montezuma	41	29	22	30
Montrose	43	35	30	33
Morgan	34	55	42	52
Otero	53	64	64	57
Ouray	14	11	7	15
Park	13	17	16	11
Phillips	58	58	53	60
Pitkin	7	3	6	3
Prowers	56	60	48	59



COUNTY	Cancer Index Rank	CVD Index Rank	Pulmonary Index Rank	HEAL Index Rank	
Pueblo	54	61	62	64	
Rio Blanco	18	32	26	13	
Rio Grande	37	51	47	46	
Routt	5	4	8	5	
Saguache	51	45	25	41	
San Juan	10	5	10	21	
San Miguel	6	1	1	29	
Sedgwick	50	52	58	51	
Summit	1	2	5	1	
Teller	38	24	40	14	
Washington	48	44	56	58	
Weld	25	43	34	40	
Yuma	29	40	41	44	



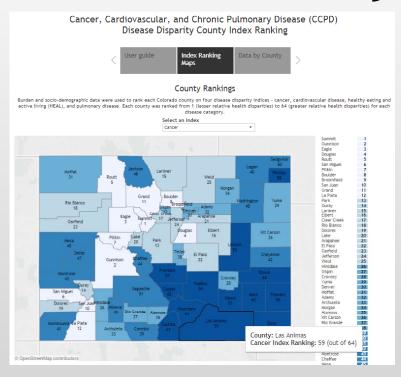
Implementation

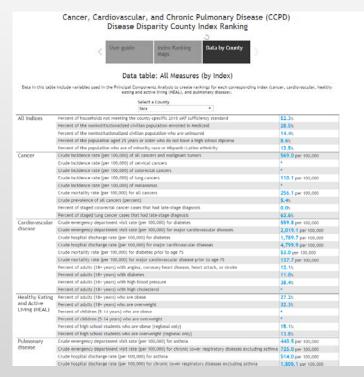
Maps, Index variables, and data distributed with the RFP by CCPD Review Committee

Up to 10 pt. award based on ranking.



Publicly Available





https://cohealthviz.dphe.state.co.us/t/HSEBPublic/views/CCPDDiseaseDisparityCountyIndexRanking



Questions?

Benjamin.White@state.co.us

303.692.6317



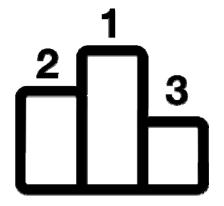
The following slides are for methodological, process documentation, and results purposes



CCPD County Index Rankings via Principal Components Analysis

Benjamin White, MPH Renee Calanan, Ph.D Elisabeth Meyer, MPH

March 2017





Examples of the use of indices using PCA

© The Author 2006. Published by Oxford University Press in association with The London School of Hygiene and Tropical Medicine. All rights reserved. doi:10.1093/heapol/czl029 Advance Access publication 9 October 2006

How to do (or not to do)...

Constructing socio-economic status indices: how to use principal components analysis

SEEMA VYAS AND LILANI KUMARANAYAKE HIVTools Research Group, Health Policy Unit, Department of Public Health and Policy, London School of Hygiene and Tropical Medicine, London, UK

"Measuring early childhood health: a composite index comparing Colombian departments"

Ana María Osório, Catalina Bolancé and Manuela Alcañiz

Journal of Urban Health: Bulletin of the New York Academy of Medicine, Vol. 83, No. 6 doi:10.1007/s11524-006-9094-x

© 2006 The New York Academy of Medicine

The Development of a Standardized Neighborhood Deprivation Index

Lynne C. Messer, Barbara A. Laraia, Jay S. Kaufman, Janet Eyster, Claudia Holzman, Jennifer Culhane, Irma Elo, Jessica G. Burke, and Patricia O'Campo

Krefis et al. Malaria Journal 2010, 9:201 http://www.malariajournal.com/content/9/1/20



RESEARCH

Open Access

Principal component analysis of socioeconomic factors and their association with malaria in children from the Ashanti Region, Ghana

Anne Caroline Krefis¹*, Norbert Georg Schwarz¹, Bernard Nkrumah³, Samuel Acquah³, Wibke Loag¹, Nimako Sarpong³, Yaw Adu-Sarkodie⁴, Ulrich Ranft², Jürgen May¹



Caveats to PCA

- Low # of observations (N=64)
 - Older literature rules suggest minimum of 200 observations,
 <100 is "poor"
 - Newer thought process involves obs/variable ratios of 2:1, 3:1 5:1...



- Missing data
 - Pairwise, Listwise, Mean/Median imputation, EM Algorithm, M.I.C.E
- Multiple ways of doing factor analysis
 - Any different way (weights, rotations) using the same data could produce very different results
- Quantifying differences in index rankings and real world utility
 - 15th is better than 16th EASY
 - 15th is N% better than 16th HARD



Variables Utilized for Indices Construction

Cancer

- All cancers combined, crude incidence rate per 100,000 persons (2013-2015, Colorado Central Cancer Registry (CCCR))
- All cancers combined, crude prevalence (%) (2013-2015, CCCR)
- All cancers combined, crude mortality rate per 100,000 persons (2013-2015, vital statistics)
- Cervical cancer, crude incidence rate per 100,000 persons (2013-2015, CCCR)
- Melanoma, crude incidence rate per 100,000 persons (2013-2015, CCCR)
- Colorectal cancer, crude incidence rate per 100,000 persons (2013-2015, CCCR)
- Lung cancer, crude incidence rate per 100,000 persons (2013-2015, CCCR)
- Colorectal cancer, % late stage diagnosis (2013-2015, CCCR)
- Lung cancer, % late stage diagnosis (2013-2015, CCCR)

CVD/Diabetes

- Percent of adults who have been diagnosed with high blood pressure (2013-2015, BRFSS)
- Percent of adults aged 20+ years who have been diagnosed with high cholesterol (2013 & 2015, BRFSS)
- Percent of adults who have been diagnosed with angina/CHD, heart attack, or stroke (2013-2015, BRFSS)
- Percent of adults who have been diagnosed with diabetes (2013-2015, BRFSS)
- Diabetes hospital discharge rate per 100,000 (Colorado Hospital Association, 2013-2015)
- Diabetes emergency department visits per 100,000 (Colorado Hospital Association, 2013-2015)
- Diabetes premature (before age 75)
- mortality rate per 100,000 (2013-2015, vital statistics)
- Major Cardiovascular Disease hospital discharge rate per 100,000 (Colorado Hospital Association, 2013-2015)
- Major Cardiovascular Disease emergency department visits per 100,000 (Colorado Hospital Association, 2013-2015)
- Major cardiovascular disease premature (before age 75) mortality rate per 100,000 (2013-2015, vital statistics)

Pulmonary

- Percent of children who currently have asthma (2013-2015, Child Health Survey)
- Percent of high school students who have ever been diagnosed with asthma (2015, Healthy Kids Colorado Survey)
- Percent of adults who currently have asthma (2013-2015, BRFSS)
- Percent of adults who have been diagnosed with COPD (2013-2015, BRFSS)
- Asthma Hospital discharge rate per 100,000 (Colorado Hospital Association, 2013-2015)
- Asthma Emergency Department Visits per 100,000 (Colorado Hospital Association, 2013-2015)
- Chronic Lower Respiratory Disease Hospital discharge rate per 100,000 (Colorado Hospital Association, 2013-2015)
- Chronic Lower Respiratory Disease Emergency Dept. Visits per 100,000 (Colorado Hospital Association, 2013-2015)
- Chronic Lower Respiratory Disease (exclusive of asthma) mortality rate per 100,000 (2013-2015, vital statistics)

HEAL

- Percent of adults who are obese (2013-2015, BRFSS)
- Percent of high school students who are obese (2015, Healthy Kids Colorado Survey (HKCS))
- Percent of children aged 5-14 who are obese (2013-2015, Child Health Survey)
- Percent of adults who are overweight (2013-2015, BRFSS)
- Percent of high school students who are overweight (2015, Healthy Kids Colorado Survey (HKCS))
- Percent of children aged 5-14 who are overweight (2013-2015, Child Health Survey)

Social Determinants of Health (used in all indices)

- Population Density, Persons per Square Mile (U.S. Census ACS 5 yr., 2011-2015)
- Percent of Population, Minority Race or Hispanic Latino (U.S. Census ACS 5 yr., 2011-2015)
- Percent of Population Whose Income is Below the Federal Poverty Level (U.S. Census ACS 5 yr., 2011-2015)
- Percent of the Population Age 25+ Without a H.S Diploma (U.S. Census ACS 5 yr., 2011-2015)
- Percent of the Total Noninstitutionalized Civilian Population that is Insured (U.S. Census ACS 5 yr., 2011-2015)
- Annual Income Self Sufficiency Standard (as Percentage of Federal Poverty Level)

 Needed for a family with 2 adults, One preschooler and one school-age child

 (2015 Colorado Self Sufficiency Report)



Testing Normal Dist. of Variables

Variable	Mean	StDev	Shapiro-Wilk	Shapiro-Wilk (α <0.05)	Normal?
VAR1	3788.3	1290	0.988412	0.8127	Normal
VAR2	1258.03	654.1	0.969993	0.1209	Normal
VAR3	484.4	210.27	0.96676	0.082	Normal
VAR4	961.854219	543.59	0.933722	0.002	NonNormal
VAR5	3353.2	2039	0.819102	0.0001	NonNormal
VAR6	1171.05	945	0.912616	0.0001	NonNormal
VAR7	895.2	786.1	0.780887	0.0001	NonNormal
VAR8	853.9	617.3	0.795856	0.0001	NonNormal
VAR9	67	7.26	0.976	0.2644	Normal
VAR10	59.4	9.05	0.9739	0.1932	Normal
VAR11	79.47	7.16	0.9393	0.0035	NonNormal
VAR12	14135.93	26206	0.54	0.0001	NonNormal
VAR13	110	33.2	0.909	0.0002	NonNormal
VAR14	31.19	14.5	0.8974	0.0001	NonNormal
VAR15	18.77	12.01	0.93947	0.0036	NonNormal
VAR16	41.62	19.16	0.964	0.0588	Normal
VAR17	375.43	49.23	0.97	0.1277	Normal
VAR18	437	875	0.527404	0.0001	NonNormal
VAR19	29.63	7.59	0.96833	0.099	Normal

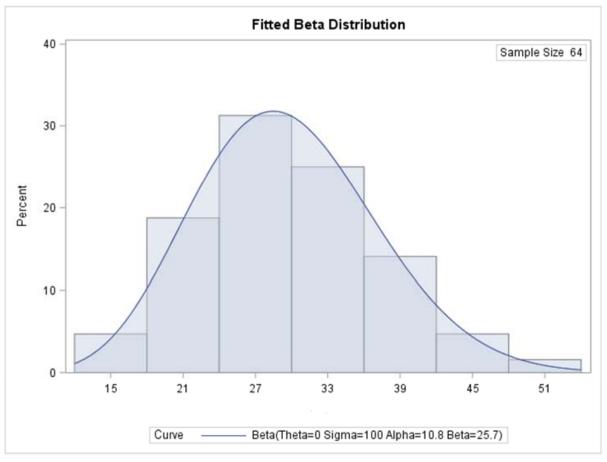
Variable	Mean	StDev	Shapiro-Wilk	Shapiro-Wilk (α <0.05)	Normal?
VAR20	36.3	6.61	0.911101	0.0002	NonNormal
VAR21	8.079	4.13	0.897	0.0001	NonNormal
VAR22	8.709	5	0.914	0.0003	NonNormal
VAR23	22066.4	43714	0.53	0.0001	NonNormal
VAR24	5.42	2.89	0.879	0.0001	NonNormal
VAR25	21.17	4.13	0.83	0.0001	NonNormal
VAR26	8.51	3.32	0.91	0.0002	NonNormal
VAR27	5.9	3.89	0.81	0.0001	NonNormal
VAR28	7467	14832	0.53	0.0001	NonNormal
VAR29	58.68	9.9	0.9822	0.4879	Normal
VAR30	23.44	6.18	0.91993	0.0005	NonNormal
VAR31	27.7	9.87	0.87	0.0001	NonNormal
VAR32	33937	68363.2	0.524	0.0001	NonNormal
VAR33	155.9	585.09	0.277	0.0001	NonNormal
VAR34	25.02	14.42	0.91857	0.0004	NonNormal
VAR35	14.38	5.74	0.95	0.0171	NonNormal
VAR36	6	2.7	0.976	0.2639	Normal
VAR37	14.38	4.49	0.969	0.108	Normal
VAR38	230.8	46.06	0.93	0.0014	NonNormal

11 (27.5%) Normal Variables, 29 (72.5%) Non-Normal Variables Many variables are non normal because they are percentage data (i.e. bounded) These data fit the Beta distribution. See next slide as example



Testing Beta Dist. of Variables:

Percent of adults who have been diagnosed with high blood pressure (2013 & 2015, BRFSS)



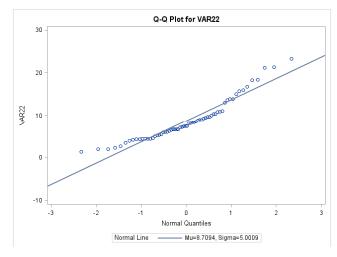


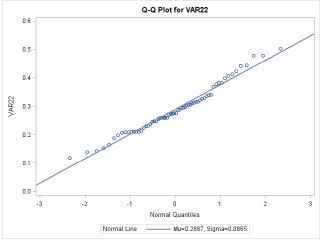
Transformation of Variables? Exploration (Arcsine)

Convert Percentage Variables (%) to TRUE Proportions (0.0 - 1.0)

Transform these true proportions using the ARCSINE transformation process.

Arcsine Transformation (along with Logit) are variance-stabilizing transformations







Transformation of Variables? NO

Variable	Mean	StDev	Shapiro-Wilk	Shapiro-Wilk (α <0.05)	Normal?	Arcsine Shapiro-Wilk	Arcsine Shapiro-Wilk (α <0.05)	Arcsine Normal?
VAR11	79.47	7.16	0.9393	0.0035	NonNormal	0.9778	0.3	Normal
VAR20	36.3	6.61	0.911101	0.0002	NonNormal	0.9217	0.0001	NonNormal
VAR21	8.079	4.13	0.897	0.0001	NonNormal	0.9509	0.0128	NonNormal
VAR22	8.709	5	0.914	0.0003	NonNormal	0.973	0.175	Normal
VAR24	5.42	2.89	0.879	0.0001	NonNormal	0.933	0.002	NonNormal
VAR25	21.17	4.13	0.83	0.0001	NonNormal	0.744	0.0001	NonNormal
VAR26	8.51	3.32	0.91	0.0002	NonNormal	0.96	0.05	Normal
VAR27	5.9	3.89	0.81	0.0001	NonNormal	0.91	0.0002	NonNormal
VAR30	23.44	6.18	0.91993	0.0005	NonNormal	0.91	0.0002	NonNormal
VAR31	27.7	9.87	0.87	0.0001	NonNormal	0.99	0.0001	NonNormal
VAR34	25.02	14.42	0.91857	0.0004	NonNormal	0.96	0.02	NonNormal
VAR35	14.38	5.74	0.95	0.0171	NonNormal	0.98	0.52	Normal

Only Four of the original non-normal proportion variables would be transformed...WHY???

Literature review suggests arcsine = arcane hahahaha ohh I've been doing this too much today.

"Common statistical assumptions of normality and homoscedasticity do not apply to PCA, which eliminates the need for data transformations that often result in a loss of original information [18]". Online J Public Health Inform. 2016; 8(2): e192.



Handling Missing Data



- Cancers (less than 3 reported redacted) -> impute 1 or 2
 OBTAINED NON-REDACTED DATA FOR THIS ANALYSIS
- Missing Data values with known neighbors replace with Health Statistics Region (HSR) value http://www.chd.dphe.state.co.us/HealthDisparitiesProfiles/dispHealthProfiles.aspx
- Missing Data values with unknown neighbors (childhood asthma & obesity, HS asthma & obesity data):
 - Stochastic iterative imputation -> Multiple Imputation by Chained Equations (MICE) or (MCMC)
 - Can be used when data violate multivariate non-normal data assumption
 - Can be used in lieu of the EM algorithm when missing data exceeds 2-3%
 - Has been proven useful when data is missing upwards of 50% (ours is 6% 26%)



Multiple Imputations of Chain Equations (MICE)



A second method available in SAS imputes missing variables using the fully conditional method (FCS) which do assume a joint distribution but instead uses a separate conditional distribution for each imputed variable. This specification may be necessary if your are imputing a variable that must only take on specific values such as a binary outcome for a logistic model or count variable for a poisson model. In simulation studies (Lee & Carlin, 2010; Van Buuren, 2007), the FCS has been show to produce estimates that are comparable to MVN method. http://stats.idre.ucla.edu/sas/seminars/multiple-imputation-in-sas/mi_new_1/

https://largescaleassessmentsineducation.springeropen.com/articles/10.1186/2196-0739-1-4

This method is widely used in practice as it handles complex missing data problems relatively easily: — Some of the benefits of the chained equations approach are that each model can be specified as desired, i.e. you can declare exactly the type of model to be used and predictors included as covariates in SAS and Stata but not IVEware which performs this type of decision making for you — This method handles arbitrary missing data patterns with categorical and continuous variables easily, widely used in practice due to ease of implementation and reliable results — Another advantage is that the variable with the least amount of missing data is imputed first and then used in subsequent imputations, then the next variable with the 2nd least amount of missing data is imputed and used in subsequent imputations, etc.. ② A disadvantage (for the statistically inclined) is lack of theoretical foundation yet results are robust and generally reliable, see Van Buuren (2007 and 2012) for more detail

http://www.isr.umich.edu/src/smp/asda/Survmeth%20616%20MI%20presentation%207%2021%202014%20final.pdf

Must look at missing patterns to asses MCAR, MAR, MNAR



Multiple Imputations of Chain Equations (MICE)



5 Imputation Iterations for the following 4 variables:

Asthma (current) - children (2013-2015), 4 counties (6.3%) Missing Asthma, ever - High school students (2015)--HSR only, 4 counties (6.3%) Missing Overweight or obese, Children aged 5-14 years (2013-2015) 17 counties (24%) Missing Overweight or obese, HS students (2015)--HSR only 4 counties (6.3%) Missing

Utilize the Predicted Mean Matching Method in SAS PROC MI. It imputes a value randomly from a set of observed values whose predicted values are closest to the predicted value for the missing value from the simulated regression model (Heitjan and Little 1991; Schenker and Taylor 1996). The predictive mean matching method ensures that imputed values are plausible and might be more appropriate than the regression method if the normality assumption is violated (Horton and Lipsitz 2001, p. 246).

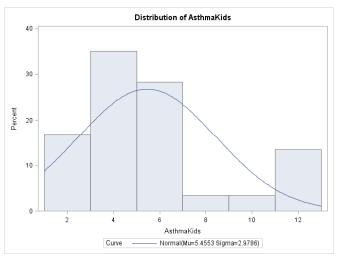
After Imputations, common practice is to Model these imputations using a regression, look at the intercepts, and then combine in a model. I didn't want to run PCA on a MODEL that has already undergone imputation, so I averaged the imputation values based on their mean and standard errors and compared them to their neighbors.

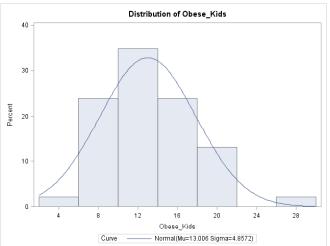
http://thestatsgeek.com/2016/03/12/combining-bootstrapping-with-multiple-imputation/ http://jeremyjaytaylor.squarespace.com/missing-dataimputation-discuss/post/1585594

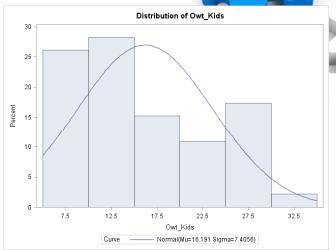


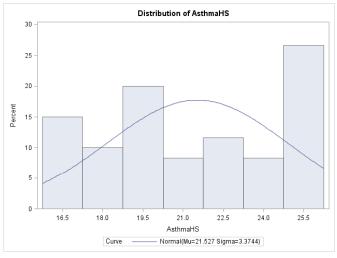
Variables for MICE

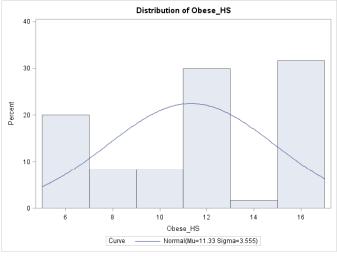


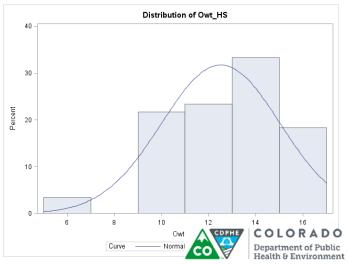












Correlated Variables for MICE Variables' Predictions



	PctHous				
	Populati Pct_Min eholds_ AllCanc AllCanc	AllCanc Diabete	Asthma		
	on_Tota Pct_Me orityRac Below_SerCrude erCrude CervCru MelaCrudel Co	oloCru Colo_Pc LungCru Lung_Pc erCrude sPrem7 MCVDPr CLRDNo MCVDHo			_ Obese_ Obese_ HBP_Ad
TYPENAME	I dicaid e SS Inc Prev delnc nc de	elnc tLate delnc tLate Mort 5 em75 Asthma sp100k	p100k k sp100k R100k 100k _ER100kR100k K	Kids HS Adults ults IAdult HAAdultsAdult ults Owt_HS ds Adults	HS Kids ults
	0.33080 0.03771 0.04044 - 0.34066416	<mark>0.20657</mark>	0.01386 -	0.07725 0.10752	<u>-</u>
CORR AsthmaKids	2 0.15649 -0.1289 0.03154 2 9 0.04742 3 0	.08608 0.05237 0.10986 3 1 0.13111 0.13423 0.15299 0.07998	8 0.13851 -0.0297 0.16366 0.05281 0.04208 8 0.09133	1 1 5 0.17708 -0.1217 0.10422 0.14238 0.00332 0.09698 <mark>0.20098 0.2502</mark>	26 0.16142 0.08836 0.13116
	- 0.43164 0.20551 0.31577 0.09877 0.21180 0.07625648 0		4 0.45372 0.35663 0.53231 0.42617 0.42441 0.47672 0		
CORR AsthmaHS	0.05108 9 0.2215 2 1 1 5 4	3 -0.015 9 0.02455 9 1 7 8 3	3 8 8 8 3 1 0.29965 6	1 1 9 8 0.14222 8 9 7 0.13056 3	4 3 0.11227 3
		0.00047	0.50000 0.0000 0.40550 0.44044 0.00507 0.40450 0.04700	0.00040.0.00000.0.00045.0.40000.0.4000	07.0.04/74
CORP. Out HE	- <mark>0.52079</mark> 0.31783 0.23977 0.19210 0.10750 - 0.16100084 0 0.16183 7 4 5 1 3 0.07859 4	.34366 - 0.23817 0.37827 0.29449 0.49714 0.53551 2 0.03038 0.35488 6 0.40032 3 2 7 8	0.53839 0.26282 0.49550 0.11211 0.20507 0.12158 0.24708	0.39012 0.36292 0.39215 0.49874 0.16182 0.50547 0.5482 0.09698 0.13056 -0.0316 1 7 3 1 2 1 3	
CORR Owt_HS	0.10183 / 4 5 1 3 0.07859 4	2 0.03038 0.35488	8 9 5 9 2 9 80	0.09098 0.13030 -0.0310 1 / 3 1 2 1 3	4 7 0.73259 4
	0.41748 0.14688 0.39555 0.19216 0.31761446 0	.35263 - 0.58502 0.09291 0.32366 0.40660 0.62821 0.69118	8 0.73368 0.56170 0.76289 0.43548 0.49760 0.41248 0.54905	- 0.57073 0.30405 0.59245 0.62646 0.46167 0.64868 0.03810 0.50547 0.7042	21 0.61649 0.54329
CORR Owt_Kids	-0.2155 0.51189 9 6 7 -0.0406 7 8	5 0.16522 4 7 0.40114 3 6 3 4		0.20098 3 8 6 9 7 4 7 3 1	5 6 0.30524 3
CORK OWI_RIGS			2 0 66903 0 38575 0 63367 0 24497 0 33729 0 23836 0 38083		
CORR Obese_HS	0.12308 3 1 1 3 3 0.04545 0.19605639	9 0.12676 0.43742 8 7 6 6 0.59418 2		0.16142 3 7 2 1 2 0.58934 0.19997 7 6	0 1 2 0
000000110		.37120 - 0.08168 0.35934 0.24017 0.34706 0.39850 0.46881			98 0.60662 0.40442
CORR Obese Kids	0.33846 0.44345 4 9 8 7 0.05644 -0.0359392			0.08836 0.11227 0.05204 3 6 5 1 0.04483 0.73259 0.30524	8 3 1 4

**** BEST CORRELATED VARIABLES TO USE IN IMPUTATION:

```
For AsthmaKids: MelaCrudeInc(0.34) Population_Total(0.33) ObeseAdults (-0.25) Pct_Uninsured(-0.22) Owt_Kids(-0.20) LungPctLate(0.20) For AsthmaHS: Owt_Kids (0.57) CLRDHosp100k (0.53) MCVDPrem75 (0.49) CLRD_ER100k (0.47) Pct_Medicaid (0.43) AsthmaAdults(0.42) For Owt_Kids: Obese_Adults (0.70) DiabetesAdult (0.65) HighCholAdult(0.62) Obese_HS (0.62) HBP_Adults (0.54) Pct_Medicaid (0.51) For Owt_HS: Obese_HS (0.91) Obese_Kids(0.73) Pct_Medicaid (0.53) DiabHosp100k (0.54) Obese_Adults (0.54) For Obese_Kids: Owt_Hs (0.73) Obese_HS (0.60) Pct_Medicaid(0.51) MCVDHosp100k (0.47) For Obese_HS: Owt_HS (0.91) DiabHosp100k(0.67) Obese_Adults (0.65) CLRDHosp100k (0.64) Pct_Medicaid (0.60);
```



Multiple Imputations of Chain Equations (MICE)



	Missing Data Patterns													
									Group Means					
Group	AsthmaKids	AsthmaHS	Owt_Kids	Owt_HS	Obese_Kids	Obese_HS	Freq	Percent	AsthmaKids	AsthmaHS	Owt_Kids	Owt_HS	Obese_Kids	Obese_HS
1	X	X	X	X	X	X	42	65.63	5.180655	20.285714	16.694576	12.921429	13.474000	12.038095
2	X	Х		X		Х	14	21.88	5.775671	24.371429		12.228571		10.700000
3	X		Х		X		4	6.25	7.217275		10.907425		8.091550	
4		X		Х		X	4	6.25		24.600000		9.200000		6.100000



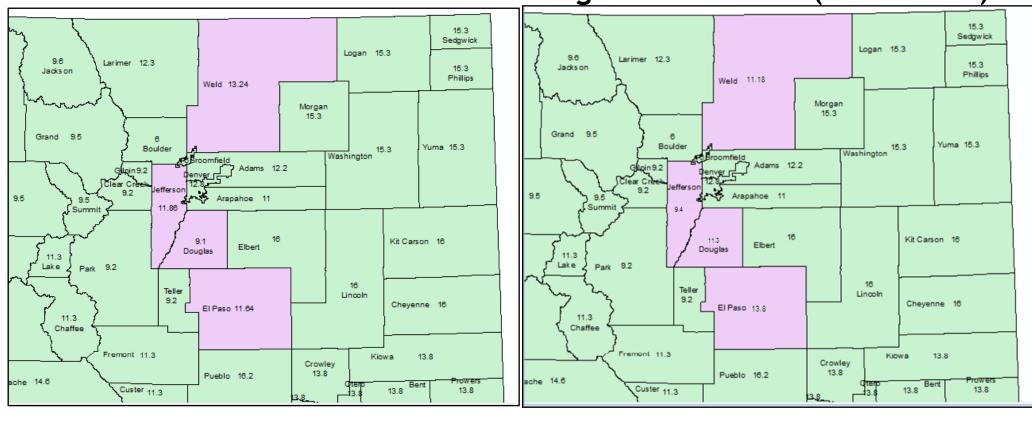
Multiple Imputations of Chain Equations (MICE), HS Overweight (%) Example



County	Variable	Imputation1	Imputation2	Imputation3	Imputation4	Imputation5	ImpAvg
Douglas	Overwt_HS	9.5	6	11.3	9.5	9.5	9.16
El Paso	Overwt_HS	11.3	11.6	13.4	12.3	9.6	11.64
Jefferson	Overwt_HS	14.6	9.2	11.3	14.6	9.6	11.86
Weld	Overwt_HS	15.3	14.6	9.6	14.6	12.1	13.24



MICE VALIDATION EXAMPLE: Overweight HS Students (2013-2015)



WICE 5 Imputation Average vs.
Known Neighbors:

Median of Neighbors

VS.

Known Neighbors

COLORADO

Department of Public Health & Environment

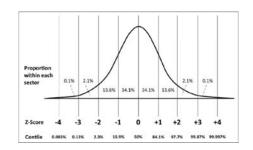
Standardizing Variables

PROC MEANS to calculate means and standard deviations for all continuous variables Create Standardized Z-scores for each continuous variable at the county level.

"Standardization is generally acknowledged as a necessary step before proceeding to an aggregation process. This is important to avoid giving variables with different measurement units and disproportionate ranges undue importance at the expense of others" (Gilthorpe, 1995).

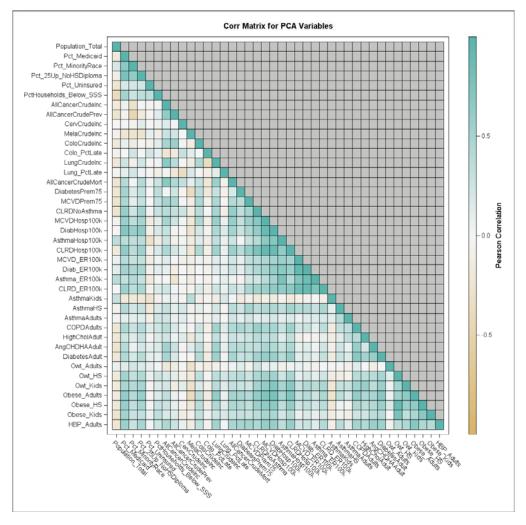
Data can be standardized even if it is non-parametric/doesn't follow a normal distribution. The question is whether or not the number of standard deviations from the mean (which is what you get when you standardize a variable) is meaningful for your data.

$$z_i = \frac{x_i - \bar{x}}{s}$$





Correlation of Standardized Variables

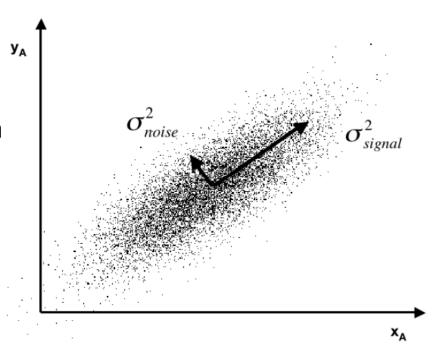




Principal Components Analysis

A type of factor analysis, which is itself a data reduction process

It filters out "noise" and reduces data down to components or "factors" of variables. These factors are the underlying structure of the data and explain the most variance in a dataset.





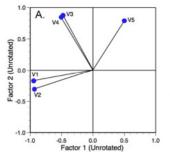
Orthogonal Varimax Rotation

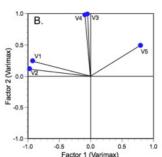
A rotation is a linear transformation that is performed on the factor solution for the purpose of making the solution easier to interpret.

A varimax rotation is an orthogonal rotation, meaning that it results in uncorrelated components. Compared to some other types of rotations, a varimax rotation tends to maximize the variance of a column of the factor pattern matrix (as opposed to a row of the matrix). This rotation is probably one of the most commonly used orthogonal rotations

http://support.sas.com/publishing/pubcat/chaps/55129.pdf

$$R_{\mathrm{VARIMAX}} = \arg\max_{R} \left(\frac{1}{p} \sum_{j=1}^k \sum_{i=1}^p (\Lambda R)_{ij}^4 - \sum_{j=1}^k \left(\frac{1}{p} \sum_{i=1}^p (\Lambda R)_{ij}^2 \right)^2 \right).$$





Rotation Considerations

Varimax -

- The most popular and one of the most common.
- Groups the variables together, makes each of your factors uncorrelated with each other.
- Allows for simplified interpretation http://www.utdallas.edu/~herve/abdi-awPCA2010.pdf
- Tends to produce group factors
- Increases "multivocality" (variables load on primary factor less, a little more on secondary factor)
- Orthogonal rotations are often criticized as being too 'simplistic' or 'artificial' because it makes are completely unrelated. In the real world we know this to hardly ever be true.
- Oblimin, Oblique also considered. Oblique the components can take any angle. Degree of correlation among factors in small

http://psych.unl.edu/psycrs/statpage/pc_rot.pdf

http://www.utdallas.edu/~herve/abdi-awPCA2010.pdf

https://jalt.org/test/PDF/Brown31.pdf



Creating the Indices

Index =
$$a_1X_1 + a_2X_2 + a_3X_3 + \dots + a_pX_p$$

where the a_i are weights to be determined from the data and the X_i are an appropriate subset of p variables measured in the survey.

"When specifying PCA, the user is given the choice of deriving eigenvectors (weights) from either the correlation matrix or the co-variance matrix of the data. If the raw data has been standardized, then PCA should use the co-variance matrix"

https://academic.oup.com/heapol/article/21/6/459/612115/Constructing-socio-economic-status-indices-how-to

I am using the total co-variance explained by each varimax-rotated factor as the weights. Other examples include using the eigenvectors, eigen values as weights, or the proportional variance



Creating the Index Considerations

While the linear aggregation method is useful when all individual indicators have the same measurement unit, provided that some mathematical properties are respected. Geometric aggregations are better suited if the modeller wants some degree of non compensability between individual indicators or dimensions. Furthermore, linear aggregations reward base-indicators proportionally to the weights, while geometric aggregations reward those countries with higher scores.

In both linear and geometric aggregations, weights express trade-offs between indicators. A deficit in one dimension can thus be offset (compensated) by a surplus in another. This implies an inconsistency between how weights are conceived (usually measuring the importance of the associated variable) and the actual meaning when geometric or linear aggregations are used. In a linear aggregation, the compensability is constant, while with geometric aggregations compensability is lower for the composite indicators with low values. In terms of policy, if compensability is admitted (as in the case of pure economic indicators), a country with low scores on one indicator will need a much higher score on the others to improve its situation when geometric aggregation is used. Thus, in benchmarking exercises, countries with low scores prefer a linear rather than a geometric aggregation. - https://www.oecd.org/std/42495745.pdf, page 33





Variables for Each Index

Initial Variables included based on a priori knowledge, data available, and recommendations from the CCPD Committee and Prevention Services Division (PSD) of CDPHE



Variables Reference for Indices

/*County The name of the county in Colorado, 64 total*/ /*Pct Medicaid Percent of County Non Institutionalized Civilians enrolled in Medicaid ACS 5 year estimates, 2011-2015*/ ACS 5 year estimates, 2011-2015*/ /*Pct MinorityRace Percent of the population who are of minority race or hispanic latino /*Pct 25Up NoHSDiploma ACS 5 year estimates, 2011-2015*/ Percent of the population age 25 and up who do not have a HS Diploma /*Pct Uninsured Percent of the Non Institutionalized Civilian population who are uninsured ACS 5 year estimates, 2011-2015*/ /*PctHouseholds Below SSS The percent of total households not meeting that county 2015 self sufficiency standard Colorado Center for Law and Policy, 2015 Report and ACS 5 year Estimates*/ /*AllCancerCrudeInc Crude incidence rate (per 100.000) of all cancers and malignant tumors, 2013-2015 Colorado Cancer Registry, CDPHE*/ /*AllCancerCrudePrev Crude prevalence rate for all cancers Colorado Cancer Registry, CDPHE*/ /*CervCrudeInc Crude incidence rate (per 100,000) of cervical cancers, 2013-2015 Colorado Cancer Registry, CDPHE*/ /*MelaCrudeInc Crude incidence rate (per 100,000) of melanomas, 2013-2015 Colorado Cancer Registry, CDPHE*/ Crude incidence rate (per 100,000) of colorectal cancers, 2013-2015 /*ColoCrudeInc Colorado Cancer Registry, CDPHE*/ /*Colo PctLate The percentage of colorectal cancers diagnosed as late stage, delayed diagnosis Colorado Cancer Registry, CDPHE*/ Colorado Cancer Registry, CDPHE*/ /*LungCrudeInc Crude incidence rate (per 100,000) of lung cancers, 2013-2015 /*Lung PctLate The percentage of lung cancers diagnosed as late stage, delayed diagnosis Colorado Cancer Registry, CDPHE*/ /*AllCancerCrudeMort Crude mortality rate (per 100,000) for all cancers, all races, 2013-2015 CDPHE Vital Statistics */ Crude mortality rate (per 100,000) for diabetes under 75yrs, all races, 2013-2015 CDPHE Vital Statistics */ /*DiabetesPrem75 /*MCVDPrem75 Crude mortality rate (per 100,000) for major cardiovascular disease events under 75yrs, all races, 2013-2015 CDPHE Vital Statistics */ /*CLRDNoAsthma Crude mortality rate (per 100,000) for chronic lower respiratory disease, all races, 2013-2016 CDPHE Vital Statistics */ /*MCVDHosp100k Crude hospitalization rate (per 100,000) for all major cardiovascular disease events, 2013-2015 Colorado Hospital Association Data*/ /*DiabHosp100k Crude hospitalization rate (per 100,000) for diabetes, 2013-2015 Colorado Hospital Association Data*/ /*AsthmaHosp100k Crude hospitalization rate (per 100,000) for asthma, 2013-2015 Colorado Hospital Association Data*/ /*CLRDHosp100k Crude hospitalization rate (per 100,000) for chronic lower respiratory disease, 2013-2015 Colorado Hospital Association Data*/ /*MCVD ER100k Crude ED visit rate (per 100,000) for all major cardiovascular disease events, 2013-2015 Colorado Hospital Association Data*/ /*Diab ER100k Crude ED visit rate (per 100.000) for diabetes, 2013-2015 Colorado Hospital Association Data*/ /*Asthma_ER100k Crude ED visit rate (per 100,000) for asthma, 2013-2015 Colorado Hospital Association Data*/ /*CLRD ER100k Crude ED visit rate (per 100,000) for chronic lower respiratory disease, 2013-2015 Colorado Hospital Association Data*/ /*AsthmaKids Percent of kids (age 5-14) with Asthma, Health Kids Colorado Survey 2013-2015 Health Kids Colorado Survey*/ /*AsthmaHS Percent of HS age kids with Asthma, Health Kids Colorado Survey 2013-2015 Health Kids Colorado Survey*/ /*AsthmaAdults Percentage of Adults 18+ with Asthma. BRFSS. 2013-2015*/ /*COPDAdults Percentage of Adults 18+ with COPD BRFSS, 2013-2015*/ Percentage of Adults 18+ with High Cholesterol BRFSS, 2013-2015*/ /*HighCholAdult /*AngCHDHAAdult Percentage of Adults 18+ with Angina, Chronic Heart Disease, or Heart Attack BRFSS. 2013-2015*/ /*DiabetesAdult Percentage of Adults 18+ with Diabetes BRFSS, 2013-2015*/ /*Owt Adults Percentage of Adults 18+ who are classified as overweight based on BMI BRFSS, 2013-2015*/ /*Owt_HS Percentage of HS age kids who are classified as overweight based on BMI Health Kids Colorado Survey*/ /*Owt_Kids Percentage of Children (5-14) who are classified as overweight based on BMI Health Kids Colorado Survey*/ /*Obese Adults Percentage of Adults 18+ who are classified as obese based on BMI BRFSS, 2013-2015*/ /*Obese HS Percentage of HS age kids who are classified as obeset based on BMI Health Kids Colorado Survev*/ /*Obese Kids Percentage of Children (5-14) who are classified as obese based on BMI Health Kids Colorado Survev*/ /*HBP Adults Percentage of Adults 18+ who are classified as having high blood pressure BRFSS, 2013-2015*/

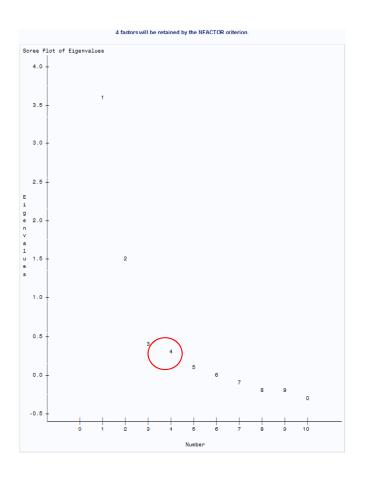




Cancer Index

Retain TWO Principal Components (based on Kaiser Criteria, EV>=1)

Eigenv	Eigenvalues of the Reduced Correlation Matrix: Total = 8.09626546 Average = 0.57830468								
	Eigenvalue	Difference	Proportion	Cumulative					
1	3.13889020	0.37121182	0.3877	0.3877					
2	2.76767838	1.61095970	0.3418	0.7295					
3	1.15671868	0.36861914	0.1429	0.8724					
4	0.70009954	0.17913154	0.0973	0.9690					
5	0.60896800	0.20269369	0.0752	1.0450					
6	0.40627432	0.25568738	0.0502	1.0952					
7	0.15058694	0.14992116	0.0186	1.1138					
8	0.00066577	0.06017004	0.0001	1.1138					
9	05950427	0.02641511	-0.0073	1.1065					
10	08591938	0.03431897	-0.0106	1.0959					
11	12023834	0.07102641	-0.0149	1.0810					
12	19126476	0.03319372	-0.0236	1.0574					
13	22445848	0.01577266	-0.0277	1.0297					
14	24023114		-0.0297	1.0000					



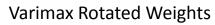




Cancer Index

Rotated Factor Pattern								
	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6		
AllCancerCrudeInc	0.83477				0.33712			
LungCrudeInc	0.72230							
AllCancerCrudePrev	0.71986	-0.32156						
ColoCrudeInc	0.71808							
AllCancerCrudeMort	0.62731	0.39461	0.33335					
Pct_25Up_NoHSDiploma		0.84419						
Pct_Medicaid		0.80376		0.37829				
Pct_MinorityRace		0.75962						
CervCrudeInc			0.66381					
Lung_PctLate			-0.56462					
PctHouseholds_Below_SSS		0.30172		0.61533				
Pct_Uninsured				0.42872				
MelaCrudeinc					0.69515			
Colo_PctLate						0.65804		
Values less than 0.3 are not p	rinted.							

	Variance Explained by Each Factor								
	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6			
<	2.7806484	2.4977429	0.9971675	0.9282451	0.8557022	0.8071230			





 With data that do not fit the common factor model perfectly, you can expect some of the eigenvalues to be negative...negative eigenvalues cause the cumulative proportion of variance explained to exceed 1 for a sufficiently large number of factors. https://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#statug_factor_sect022.htm

 "The common factor model postulates that each indicator in a set of observed measures is a linear function of one or more common

factors and one unique factor.

http://www.ophi.org.uk/wp-content/uploads/Brown-Ch-2.pdf

Eigenvalues of the Reduced Correlation Matrix: Total = 5.21644798 Average = 0.5216448								
	Eigenvalue	Difference	Proportion	Cumulative				
1	3.60024749	2.08584667	0.6902	0.6902				
2	1.51440082	1.14815265	0.2903	0.9805				
3	0.36624817	0.10453714	0.0702	1.0507	????			
4	0.26171104	0.15717760	0.0502	1.1009	2222			
5	0.10453343	0.09415651	0.0200	1.1209	2222			
6	0.01037692	0.06467905	0.0020	1.1229	????			
7	05430212	0.10075096	-0.0104	1.1125	????			
8	15505308	0.01139952	-0.0297	1.0828	????			
9	16645260	0.09880950	-0.0319	1.0509	????			
10	26526210		-0.0509	1.0000				



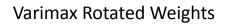
Cancer Index Output

	Rotated Factor Pattern								
		Factor1	Factor2	Factor3	Factor4				
VAR12	VAR12	0.73348							
VAR33	VAR33	0.60404	-	-					
VAR9	VAR9	0.56616	-	0.38031					
VAR11	VAR11	0.49622	-0.38558	0.41484					
VAR36	VAR36		0.83559	-					
VAR34	VAR34		0.67477						
VAR38	VAR38	0.48574	-0.63027		0.31094				
VAR37	VAR37			-0.47937					
VAR10	VAR10	0.33609			0.54402				
Values	Values less than 0.3 are not printed.								

 Variance Explained by Each Factor

 Factor1
 Factor2
 Factor3
 Factor4

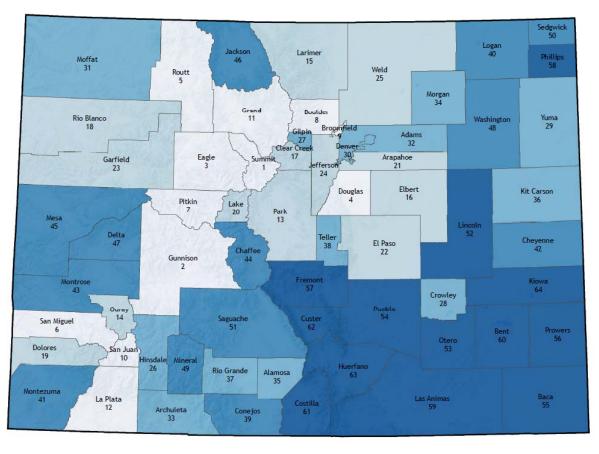
 1.9213130
 1.8484646
 0.7239138
 0.5153233







Cancer Screening Index Rankings









Cardiovascular Disease Index Variables

CVD/Diabetes

- Percent of adults who have been diagnosed with high blood pressure (2013-2015, BRFSS)
- Percent of adults aged 20+ years who have been diagnosed with high cholesterol (2013 & 2015, BRFSS)
- Percent of adults who have been diagnosed with angina/CHD, heart attack, or stroke (2013-2015, BRFSS)
- Percent of adults who have been diagnosed with diabetes (2013-2015, BRFSS)
- Diabetes hospital discharge rate per 100,000 (Colorado Hospital Association, 2013-2015)
- Diabetes emergency department visits per 100,000 (Colorado Hospital Association, 2013-2015)
- Diabetes premature (before age 75)
- mortality rate per 100,000 (2013-2015, vital statistics)
- Major Cardiovascular Disease hospital discharge rate per 100,000 (Colorado Hospital Association, 2013-2015)
- Major Cardiovascular Disease emergency department visits per 100,000 (Colorado Hospital Association, 2013-2015)
- Major cardiovascular disease premature (before age 75) mortality rate per 100,000 (2013-2015, vital statistics)

Social Determinants of Health (used in all indices)

- Population Density, Persons per Square Mile (U.S. Census ACS 5 yr., 2011-2015)
- Percent of Population, Minority Race or Hispanic Latino (U.S. Census ACS 5 yr., 2011-2015)
- Percent of Population Whose Income is Below the Federal Poverty Level (U.S. Census ACS 5 yr., 2011-2015)
- Percent of the Population Age 25+ Without a H.S Diploma (U.S. Census ACS 5 yr., 2011-2015)
- Percent of the Total Noninstitutionalized Civilian Population that is Insured (U.S. Census ACS 5 yr., 2011-2015)
- Annual Income Self Sufficiency Standard (as Percentage of Federal Poverty Level) Needed for a family with 2 adults,
 One preschooler and one school-age child (2015 Colorado Self Sufficiency Report)

Obs/Vars Ratio: 3.7:1

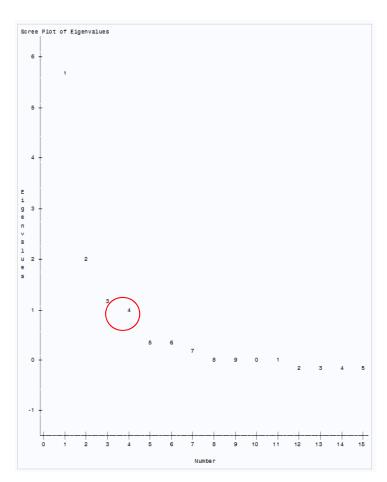




Cardiovascular Disease Index

Retain FOUR Principal Components (based on Kaiser Criteria, EV>=1)

Eigen	values of the Reduced	d Correlation Matrix:	Total = 10.1524022 A	verage = 0.67682681
	Eigenvalue	Difference	Proportion	Cumulative
1	6.02841088	4.27554577	0.5938	0.5938
2	1.75286511	0.44371147	0.1727	0.7664
3	1.30915365	0.59726239	0.1290	0.8954
4	0.71189126	0.28734315	0.0701	0.9655
5	0.42454811	0.15309931	0.0418	1.0073
6	0.27144880	0.04928932	0.0267	1.0341
7	0.22215948	0.18698386	0.0219	1.0560
8	0.03517562	0.02713863	0.0035	1.0594
9	0.00803699	0.02781926	0.0008	1.0602
10	01978228	0.02358653	-0.0019	1.0583
11	04336880	0.02248210	-0.0043	1.0540
12	06585091	0.03459045	-0.0065	1.0475
13	10044136	0.06189378	-0.0099	1.0376
14	16233514	0.05717409	-0.0160	1.0216
15	21950923		-0.0216	1.0000



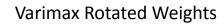




Cardiovascular Disease Index Output

Rotated Factor Pattern									
		Factor1	Factor2	Factor3	Factor4	Factor5			
DiabetesAdult		0.80727							
AngCHDHAAdult		0.69967							
MCVDHosp100k		0.67988	0.34640	0.37282		-0.33334			
HBP_Adults	HBP_Adults	0.62489			0.46155				
DiabHosp100k		0.60044	0.59892	0.39942					
HighCholAdult		0.46452							
Pct_MinorityRace			0.77402						
Pct_25Up_NoHSDiploma			0.71255						
Pct_Medicaid		0.30505	0.62705		0.52960				
MCVD_ER100k				0.92743					
Diab_ER100k			0.34201	0.89196					
DiabetesPrem75					0.60193				
MCVDPrem75		0.33921			0.56674				
PctHouseholds_Below_SSS					0.54691	0.34633			
Pct_Uninsured						0.60451			

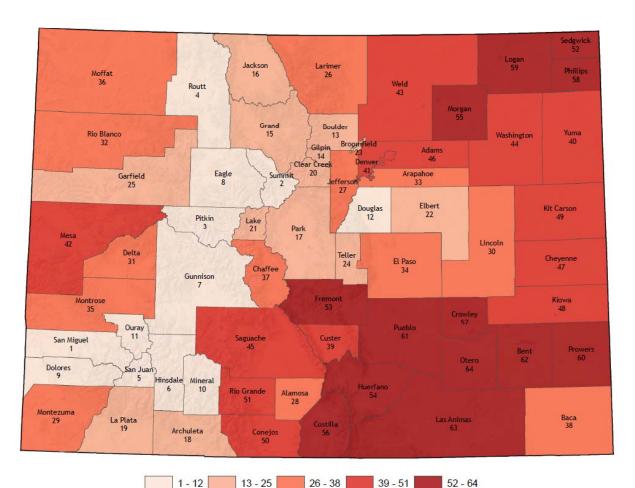
	Variance Explained by Each Factor								
	Factor1	Factor2	Factor3	Factor4	Factor5				
	2.9457860	2.3855132	2.2795200	1.7679874	0.8480625				
\									







Cardiovascular Disease Index Rankings







Pulmonary Disease Index Variables

Pulmonary

- Percent of children who currently have asthma (2013-2015, CHS)
- Percent of high school students who have ever been diagnosed with asthma (2015, HKCS)
- Percent of adults who currently have asthma (2013-2015, BRFSS)
- Percent of adults who have been diagnosed with COPD (2013-2015, BRFSS)
- Estimated number of adults who currently have asthma or have ever been diagnosed with COPD (2013-2015, BRFSS)
- Crude county rate (per 100,000) of Asthma Hospitalizations (Colorado Hospital Association, 2013-2015)
- Crude county rate (per 100,000) of Asthma Emergency Dept. Visits (Colorado Hospital Association, 2013-2015)
- Crude county rate (per 100,000) of Chronic Lower Respiratory Disease Hospitalizations (Colorado Hospital Association, 2013-2015)
- Crude county rate (per 100,000) of Chronic Lower Respiratory Disease Emergency Department Visits (Colorado Hospital Association, 2013-2015)

Social Determinants of Health (used in all indices)

- Population Density, Persons per Square Mile (U.S. Census ACS 5 yr., 2011-2015)
- Percent of Population, Minority Race or Hispanic Latino (U.S. Census ACS 5 yr., 2011-2015)
- Percent of Population Whose Income is Below the Federal Poverty Level (U.S. Census ACS 5 yr., 2011-2015)
- Percent of the Population Age 25+ Without a H.S Diploma (U.S. Census ACS 5 yr., 2011-2015)
- Percent of the Total Noninstitutionalized Civilian Population that is Insured (U.S. Census ACS 5 yr., 2011-2015)
- Annual Income Self Sufficiency Standard (as Percentage of Federal Poverty Level) Needed for a family with 2 adults,
 One preschooler and one school-age child (2015 Colorado Self Sufficiency Report)

Obs/Vars Ratio: 4.3:1

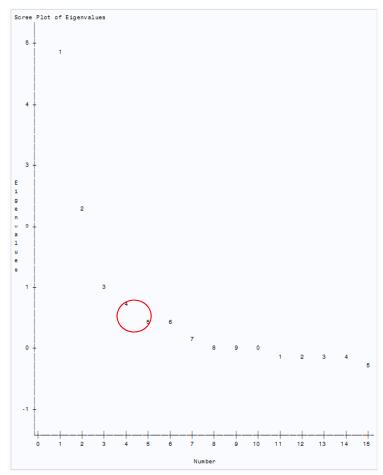




Pulmonary Disease Index

Retain THREE Principal Components (based on Kaiser Criteria, EV>=1)

Eigen	values of the Reduced	Correlation Matrix:	Total = 8.63165647 A	verage = 0.61654689
	Eigenvalue	Difference	Proportion	Cumulative
1	5.27191738	3.54958860	0.6108	0.6108
2	1.72232878	0.78018558	0.1995	0.8103
3	0.94214320	0.24732077	0.1091	0.9195
4	0.69482242	0.36880077	0.0805	0.9999
5	0.32602165	0.05114447	0.0378	1.0377
6	0.27487719	0.20190639	0.0318	1.0696
7	0.07297080	0.06248424	0.0085	1.0780
8	0.01048656	0.05850895	0.0012	1.0792
9	04802239	0.00824881	-0.0056	1.0737
10	05627121	0.03297786	-0.0065	1.0672
11	08924906	0.03237017	-0.0103	1.0568
12	12161923	0.01802042	-0.0141	1.0427
13	13963965	0.08947031	-0.0162	1.0265
14	22910996		-0.0265	1.0000







Pulmonary Disease Index

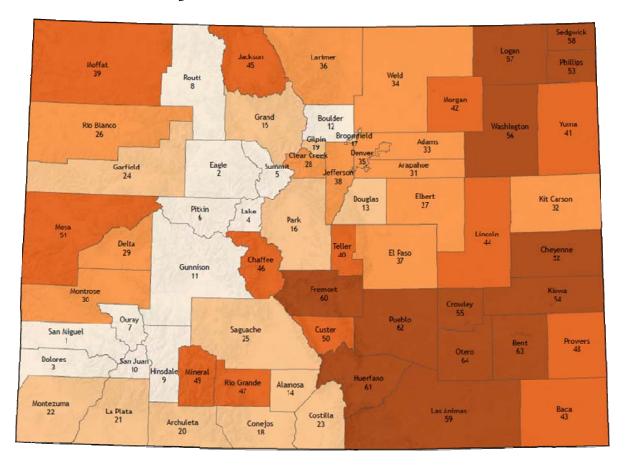
Rotated Factor Pattern									
	Factor1	Factor2	Factor3	Factor4	Factor5				
CLRDNoAsthma	0.81660								
CLRDHosp100k	0.79753	0.35839							
COPDAdults	0.63583								
Asthma_ER100k		0.91159							
CLRD_ER100k	0.43751	0.73306							
AsthmaHosp100k	0.34997	0.70180		-0.37469					
Pct_MinorityRace			0.71828						
Pct_Medicaid	0.49017		0.70999						
Pct_25Up_NoHSDiploma	0.40768		0.68408						
Pct_Uninsured				0.60951					
PctHouseholds_Below_SSS				0.56294					
AsthmaAdults					0.62839				
AsthmaHS	0.37970				0.52157				
AsthmaKids					0.31437				
Values less than 0.3 are not p	Values less than 0.3 are not printed.								

	Variance Explained by Each Factor					
/arimax Rotated Weights	Factor1	Factor2	Factor3	Factor4	Factor5	
Allillax Rotated Weights	2.6778170	2.2675943	1.9379171	1.1156257	0.9582794	





Pulmonary Disease Index Rankings









H.E.A.L. Index Variables

HEAL

- Percent of adults who are overweight or obese (2013-2015, BRFSS)
- Percent of high school students who are overweight or obese (2015, HKCS)
- Percent of children aged 5-14 who are overweight or obese (2013-2015, CHS)
- Estimated number of adults who are overweight or obese (2013-2015, BRFSS)

Social Determinants of Health (used in all indices)

- Population Density, Persons per Square Mile (U.S. Census ACS 5 yr. , 2011-2015)
- Percent of Population, Minority Race or Hispanic Latino (U.S. Census ACS 5 yr., 2011-2015)
- Percent of Population Whose Income is Below the Federal Poverty Level (U.S. Census ACS 5 yr., 2011-2015)
- Percent of the Population Age 25+ Without a H.S Diploma (U.S. Census ACS 5 yr., 2011-2015)
- Percent of the Total Noninstitutionalized Civilian Population that is Insured (U.S. Census ACS 5 yr., 2011-2015)
- Annual Income Self Sufficiency Standard (as Percentage of Federal Poverty Level) Needed for a family with 2 adults,
 One preschooler and one school-age child (2015 Colorado Self Sufficiency Report)

Obs/Vars Ratio: 6.4:1

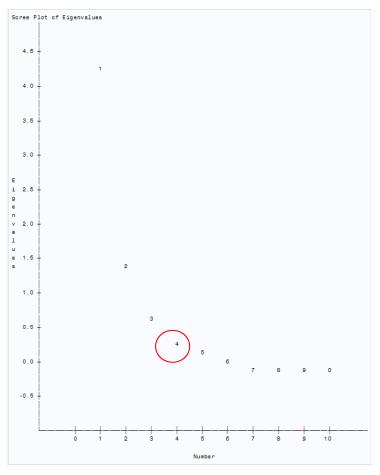




H.E.A.L. Index

Retain TWO Principal Components (based on Kaiser Criteria, EV>=1)

Eigenvalues of the Reduced Correlation Matrix: Total = 6.56551138 Average = 0.59686467						
	Eigenvalue	Difference	Proportion	Cumulative		
1	4.84120735	3.74883693	0.7374	0.7374		
2	1.09237043	0.40958935	0.1664	0.9037		
3	0.68278108	0.39644167	0.1040	1.0077		
4	0.28633941	0.17448567	0.0436	1.0514		
5	0.11185374	0.07119337	0.0170	1.0684		
6	0.04066038	0.04552804	0.0062	1.0746		
7	00486766	0.05022762	-0.0007	1.0738		
8	05509528	0.04529235	-0.0084	1.0655		
9	10038762	0.02559892	-0.0153	1.0502		
10	12598655	0.07737737	-0.0192	1.0310		
11	20336391		-0.0310	1.0000		







H.E.A.L Index

Rotated Factor Pattern					
		Factor1	Factor2	Factor3	Factor4
Pct_25Up_NoHSDiploma		0.72973		0.34893	
Obese_Adults	Obese_Adults	0.71226	0.42921		
Pct_Medicaid		0.70217	0.32743	0.48696	
Owt_Kids	Owt_Kids	0.64705	0.36308		
Pct_MinorityRace		0.58980		0.33462	0.31286
Owt_HS	Owt_HS		0.90642		
Obese_HS	Obese_HS	0.40998	0.83592		
Obese_Kids	Obese_Kids		0.76772		
PctHouseholds_Below_SSS				0.62971	
Pct_Uninsured				0.50756	
Owt_Adults	Owt_Adults				0.41161
Values less than 0.3 are not p	orinted.				

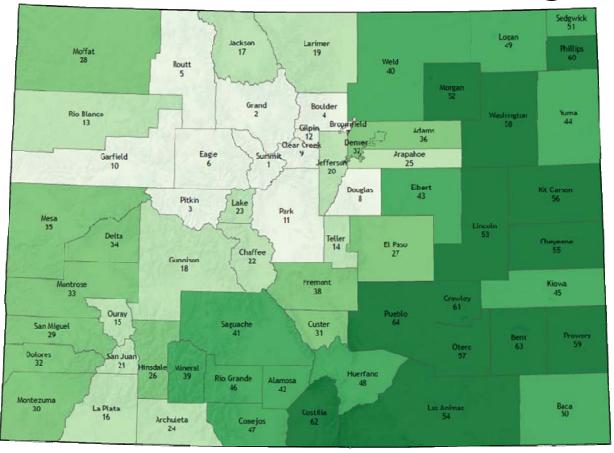
Varimax Rotated Weights	

Variance Explained by Each Factor				
Factor1	Factor2	Factor3	Factor4	
2.6214389	2.6199636	1.2135164	0.4477793	





H.E.A.L. Index Rankings







Possible Changes



- Different type of factor analysis?
- Different variables, omitting any that utilize HSR, are missing, or utilize an imputation method
- Oblique rotation vs. varimax rotation
- Different principal component weighting schema including
 - Utilizing only the first principal component, regardless of % variance explained
 - Product vs additive aggregation of principal components, utilizing perhaps a Cobb-Douglas equation:
 - new index = (PC1)^x1*(PC2)^x2 etc. where x1, x2 ... are the proportion of variance explained by that PC as a fraction of total variance explained by all https://stats.stackexchange.com/questions/133492/creating-a-single-index-from-several-principal-components-or-factors-retained-fr



Questions?

Benjamin.white@state.co.us

303.692.6317

