The University of Iowa

College of Public Health

Creating County Health Rankings for Iowa

Roger Netherton **Félix Pabón-Rodríguez**

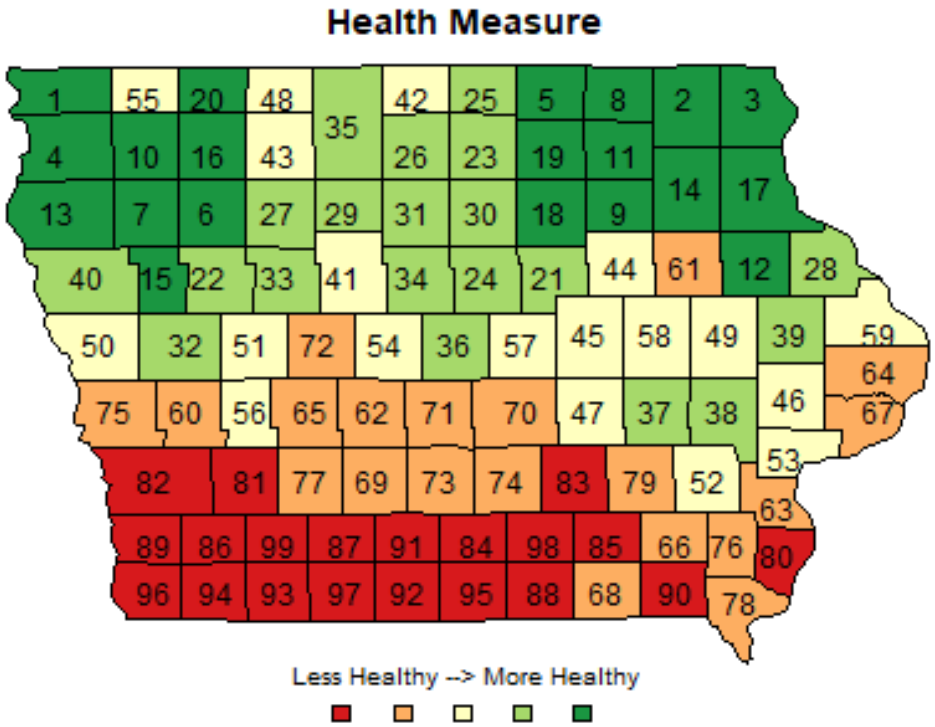
The University of Iowa
College of Public Health
Department of Biostatistics

Mentor: Jacob Oleson, Ph.D.
Iowa Summer Institute in Biostatistics (2017)

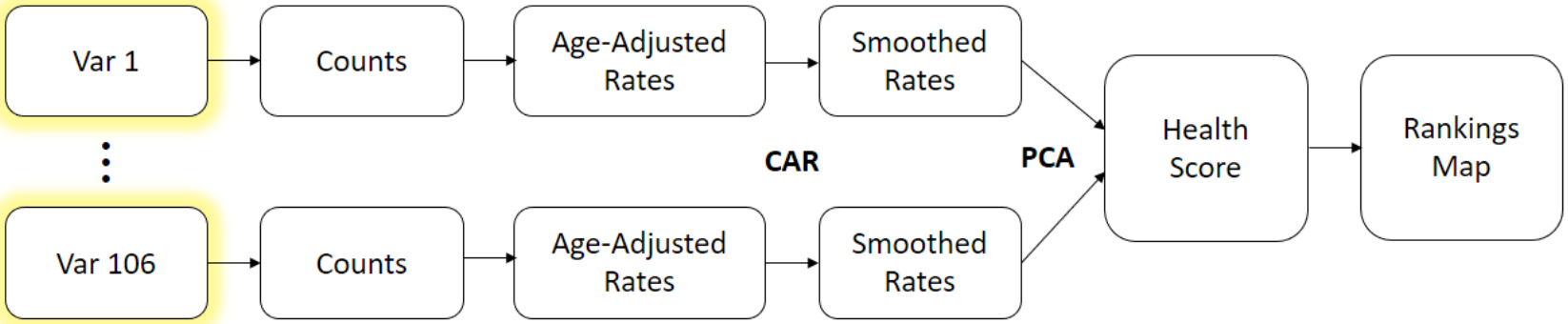
1. Objectives

- To create and visualize health rankings for the counties in Iowa based on combinations of various health measures.
- "The goal of the project is to examine where counties could use more resources to improve their overall health."

– Jake Oleson, Ph.D.

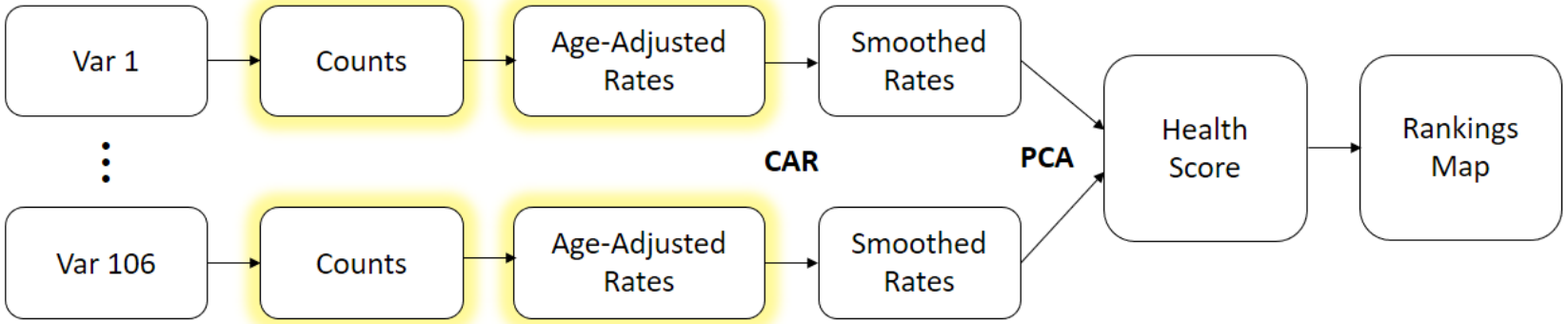


2. Dataset



- **Dataset**: Factbook 2016 data (v2.0)
- **Resource**: Iowa Health Fact Book
- **Link**: http://iowahealthfactbook.org/
- **Dimension**:
  a. 99 counties
  b. 116 health variables
  c. 76 types of measurements (counts, percents, rates, etc.)
     i. Age-Adjusted Rate
     ii. Crude Rate

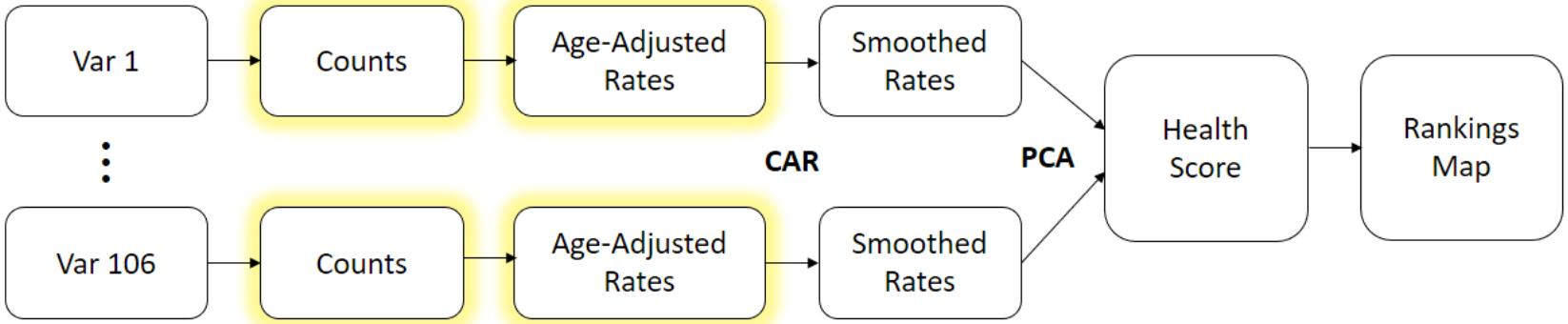
3. Definitions



Definition

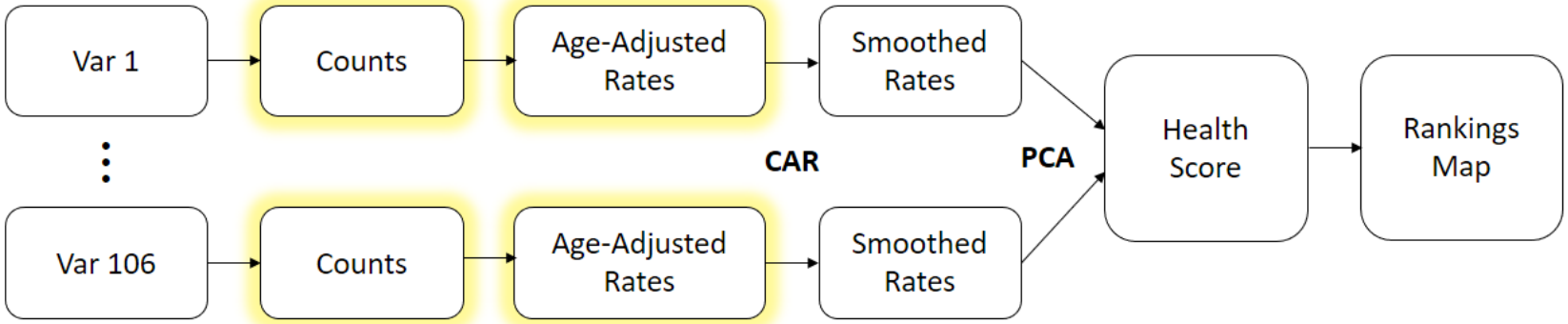
1. **Crude rate:** *the number of cases to expect if the county had a population of 100,000.*
2. **Age-adjusted rate:** *the number of cases to expect if the county had a population of 100,000 and all the counties had the same age distribution.*

4. Dataset: Sections and Subsections



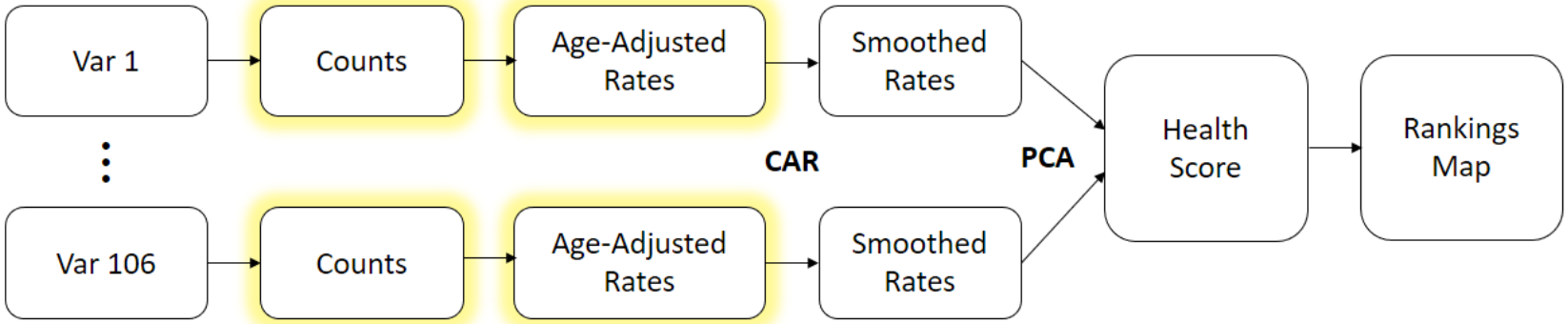
- **Cancer**
  - Colorectal Cancer Incidence, Female Breast Cancer Mortality, Leukemia Cancer Incidence, Lung Cancer Mortality, etc.
- **Social Determinants of Health**
  - Dissolutions, Marriages, Poverty Level and Underinsured
- **Health Care Facilities**
  - Hospitals, Hospital Beds, Nursing Facilities, Psychiatric Medical Institutions, Residential Care Facilities, etc.
- **Health Care Providers**
  - Dentists, Family Practice Physicians, Paramedics, Pediatric Physicians, etc.

5. Dataset: Sections and Subsections



- **Injury Mortality**
  - Transportation Mortality, Falls Mortality, Motor Vehicle Traffic Mortality and Suicides
- **Other Mortality**
  - Accidents and Adverse Effects Mortality, Alzheimer Disease Mortality, Heart Disease Mortality, Stroke Mortality, etc.
- **Prenatal and Infant Health**
  - Births, Low Birth Weight, Out of Wedlock Births, Congenital and Inherited Disorders, etc.

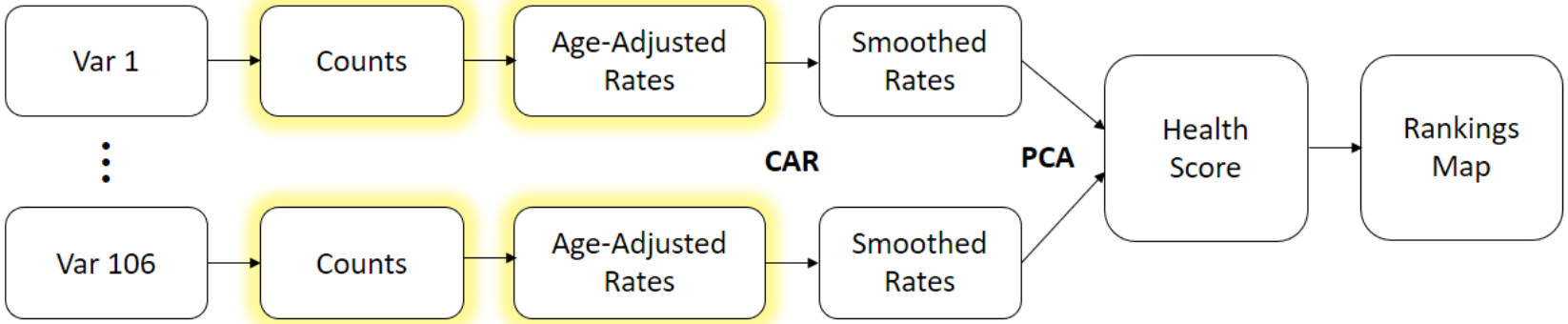
6. Outline of Code: Load & Clean



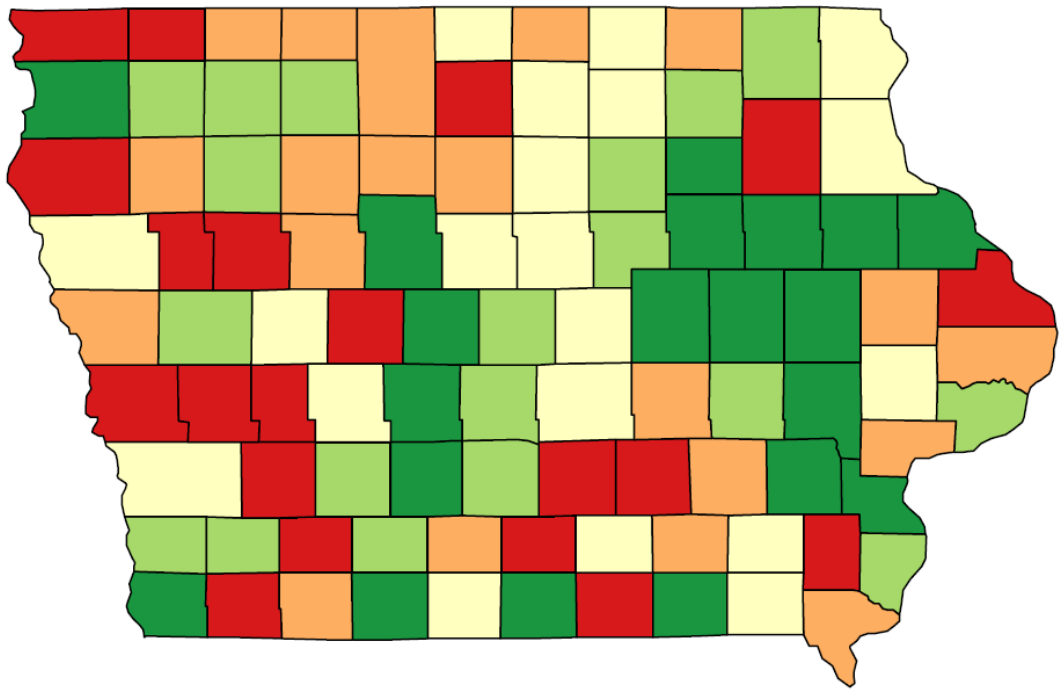
1. Load the libraries and set working directory
2. Load and clean the data
3. Fill in unobserved values (less than 5 or 10) with estimates

	County	All Cancer Incidence	All Cancer Mortality	Cervical Cancer Incidence	Cervical Cancer Mortality	Colore Cancer Incider
1	Adair	0.93187943	-0.69168958	-1.34002288	-1.32739436	0.4557
2	Adams	-2.27298515	-0.26813034	-4.06433269	-3.99131034	-2.7777
3	Allamakee	1.90740596	1.34250941	0.04338768	0.02534808	0.2703
4	Appanoose	0.26521867	-0.53564144	-0.12695230	-0.14121572	-1.0624
5	Audubon	0.75451097	-0.58022662	-2.18509953	-2.15373694	-1.1667
6	Benton	0.59854905	0.18329464	0.75342836	0.71964823	1.5219

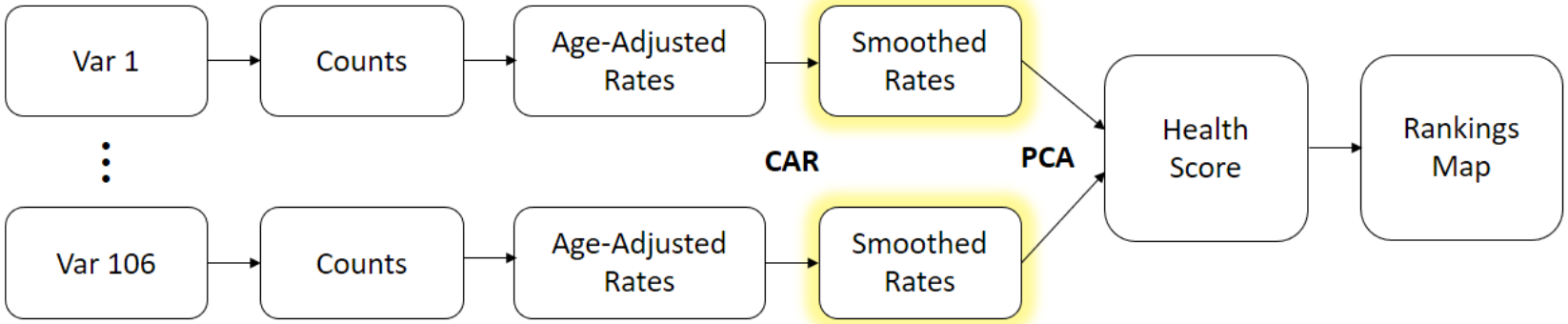
7. Outline of Code: Load & Clean



Pre-Smoothed Data



8. Method: Bayesian CAR Model

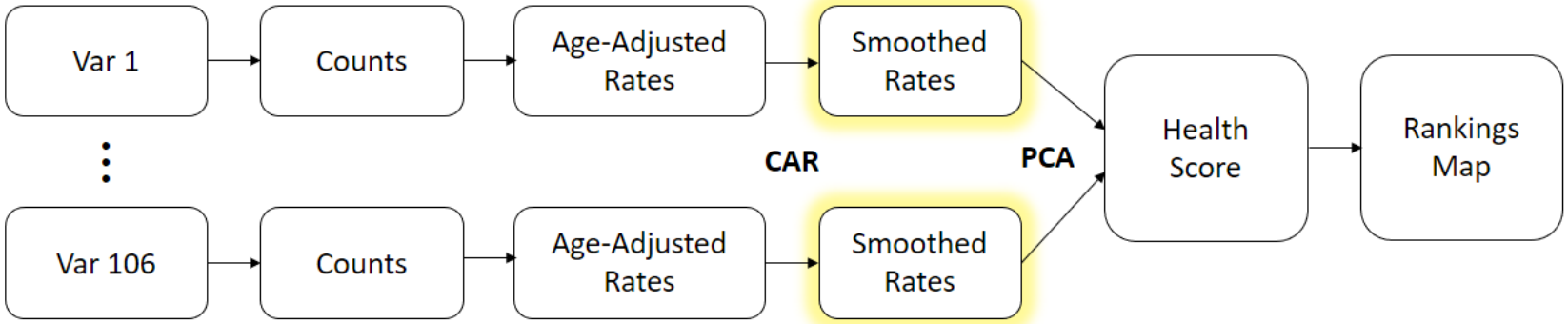


$$y_i \sim N(\mu_i, \sigma^2)$$

$$\mu_i = \beta_0 + Z_i$$

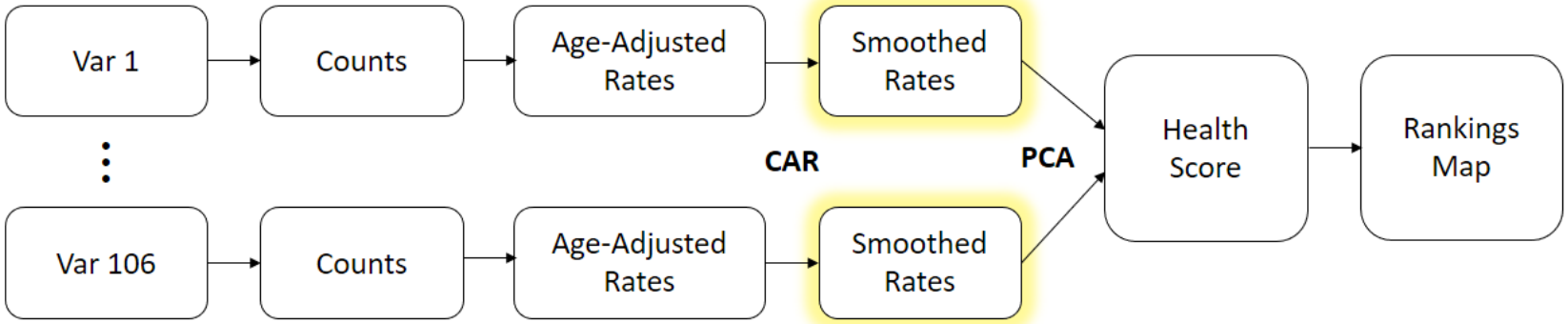
$$Z \sim N(0, \tau(I - W)^{-1}))$$

9. Outline of Code: Spatial Smoothing

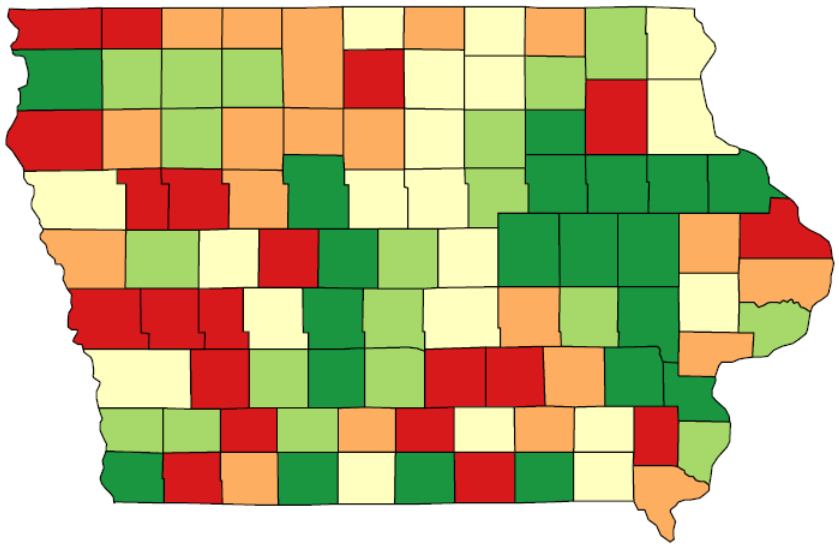


4. Smooth the values using Bayesian conditional autoregressive (CAR) model
   a. Use OpenBUGS to run the model using Markov Chain Monte Carlo (MCMC) simulation
   b. Get the estimated means of the posterior distributions for each county's health measure value
   c. Use these estimated means as the new smoothed values for each county

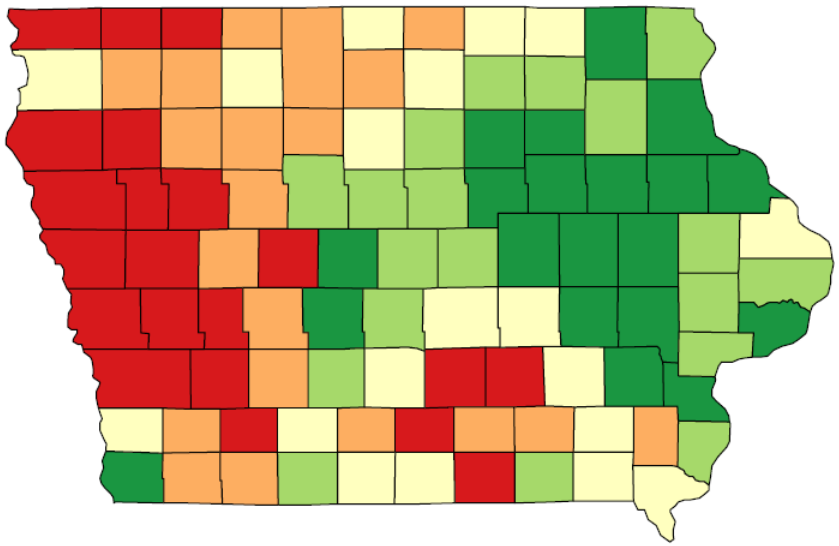
10. Outline of Code: Spatial Smoothing



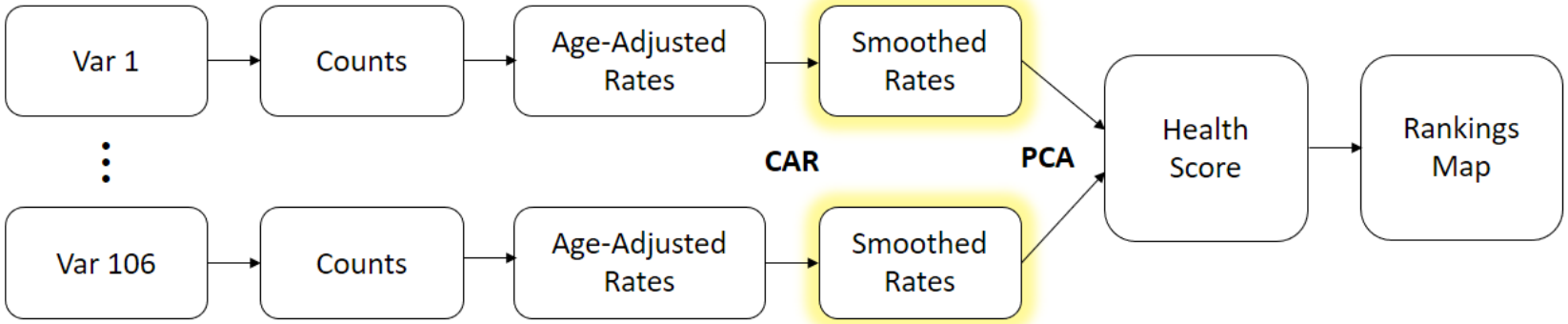
Raw Data



Smoothed Data



11. Outline of Code: OpenBUGS

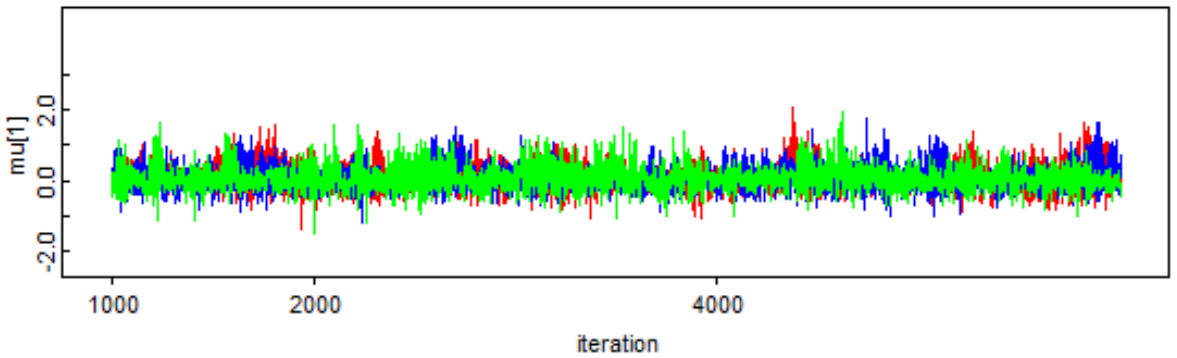


OpenBUGS - [Log]

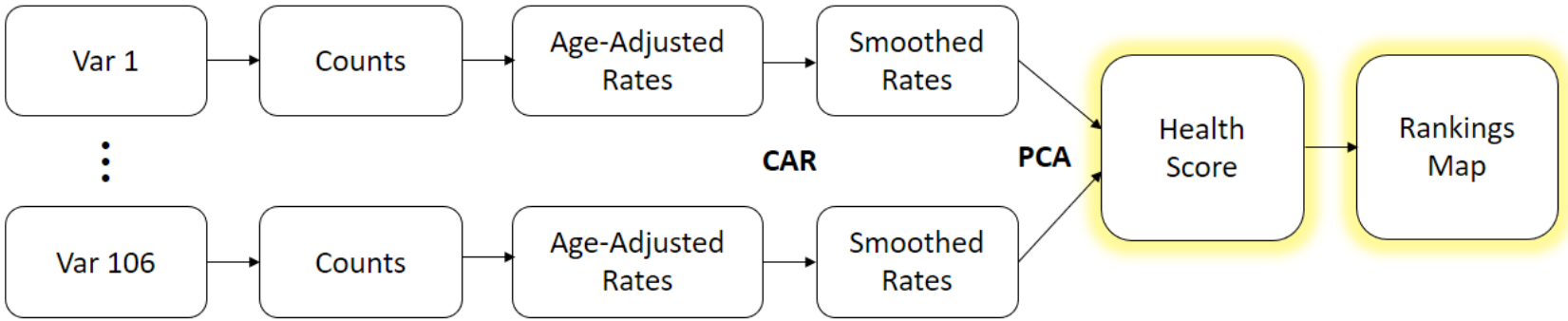
File Edit Attributes Tools Info Model Inference Doodle Map Text Window Examples Mar

Summary statistics

	mean	sd	val2.5pc	median	val97.5pc	sample
mu[1]	0.09523	0.3289	-0.4841	0.06544	0.8524	15000
mu[2]	-0.2783	0.3784	-1.16	-0.2229	0.3204	15000
mu[3]	0.5196	0.5375	-0.3	0.4266	1.773	15000
mu[4]	0.1726	0.3626	-0.4914	0.1432	0.9763	15000
. . .						
. . .						
. . .						
mu[97]	-0.1231	0.3556	-0.8987	-0.1039	0.5518	15000
mu[98]	0.08438	0.3505	-0.6108	0.07778	0.8195	15000
mu[99]	0.1084	0.2964	-0.4396	0.08928	0.7594	15000



12. Method: Principal Components Analysis

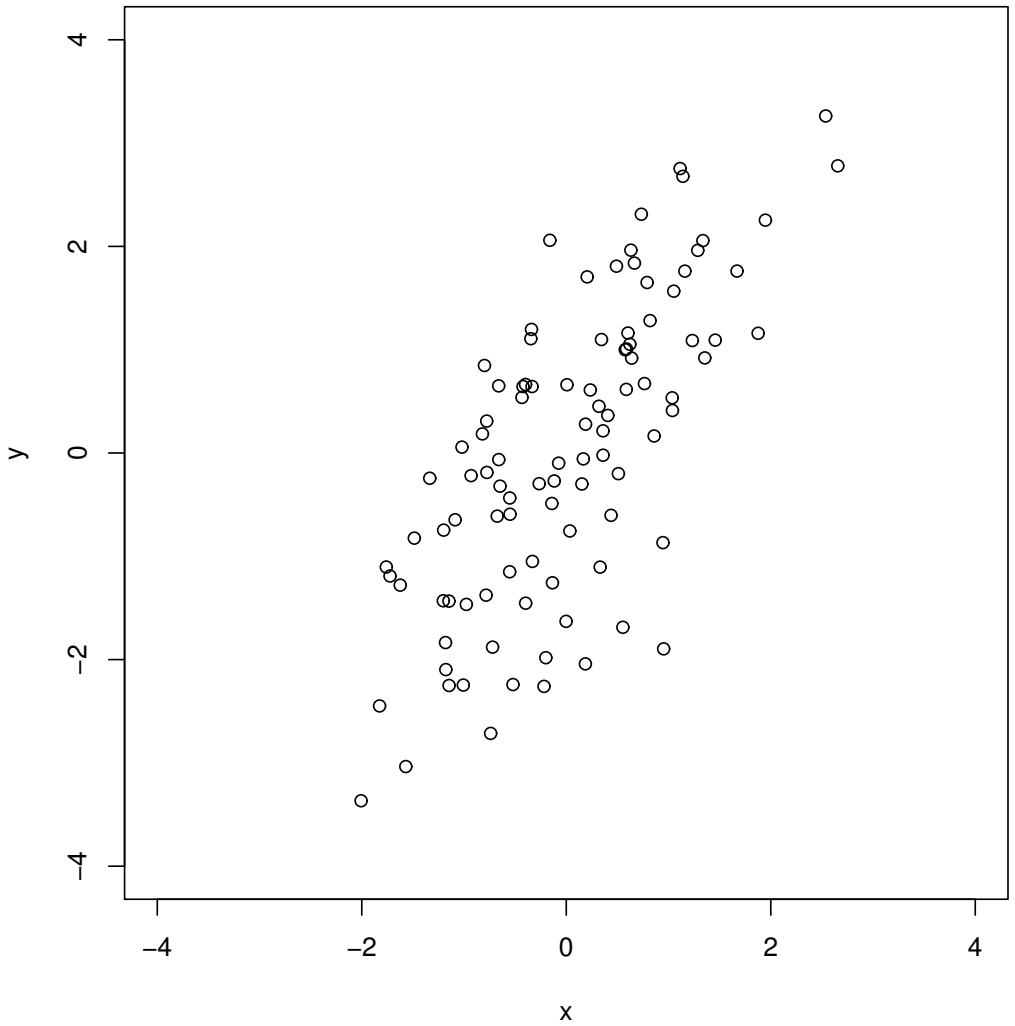


Definition

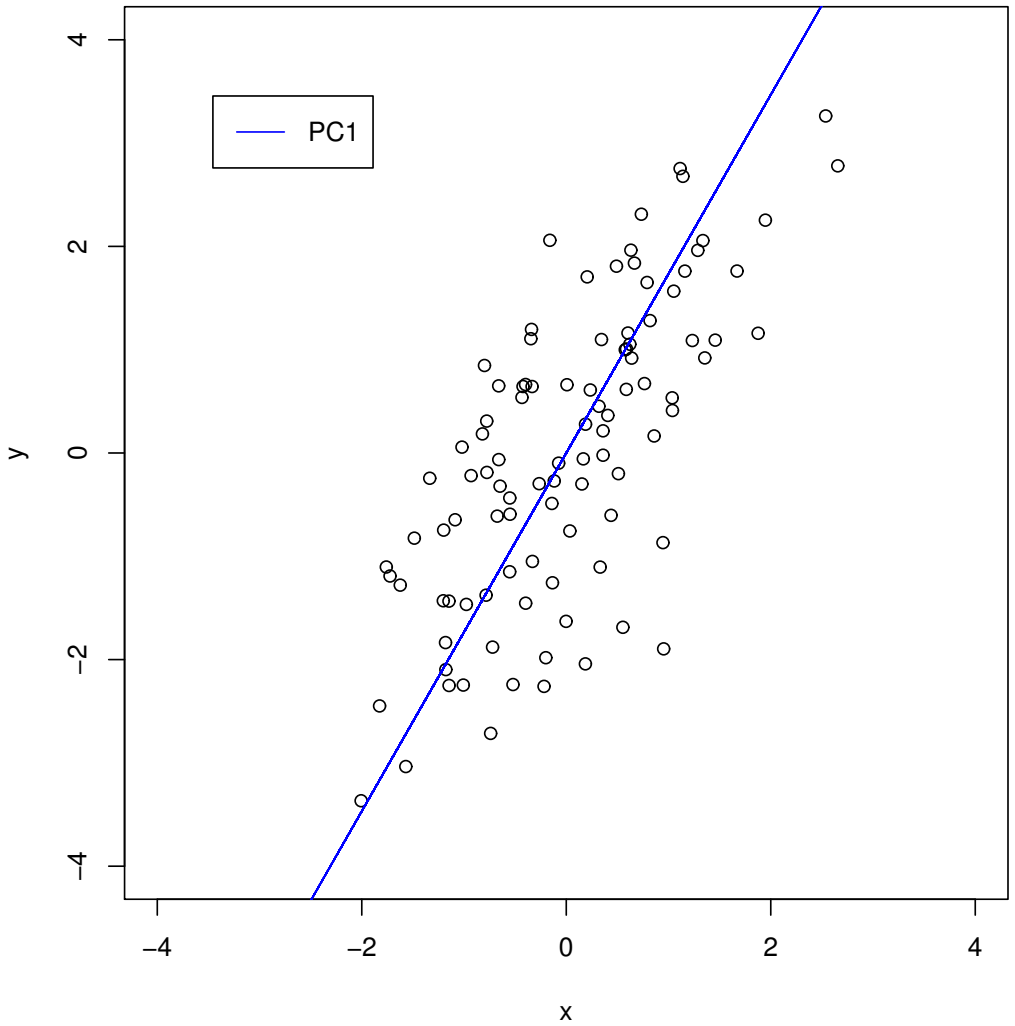
Principal Components Analysis (PCA)

- *Principal components analysis is a mathematical algorithm that reduces the dimensionality of the data while retaining most of the variation in the data set. (Ringner, 2008)*

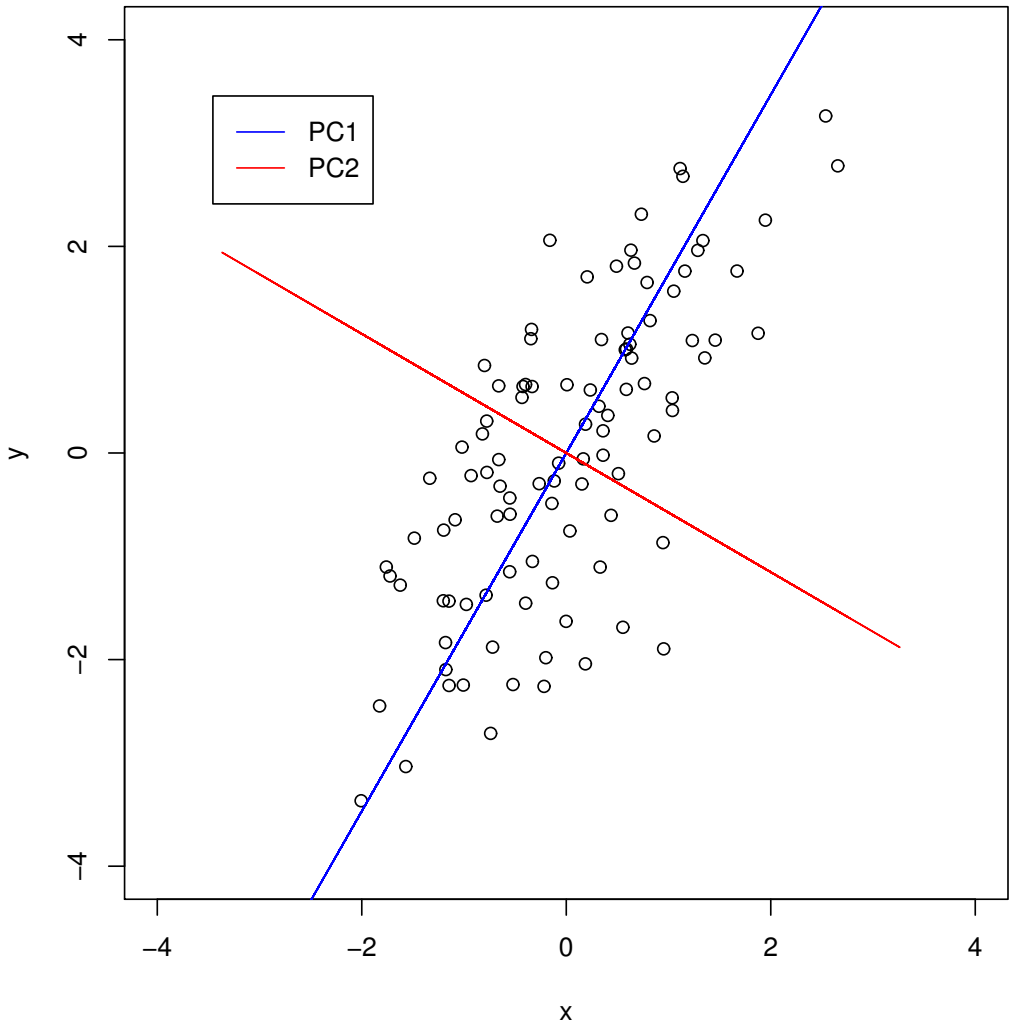
13. Example #1: Good Candidate for PCA



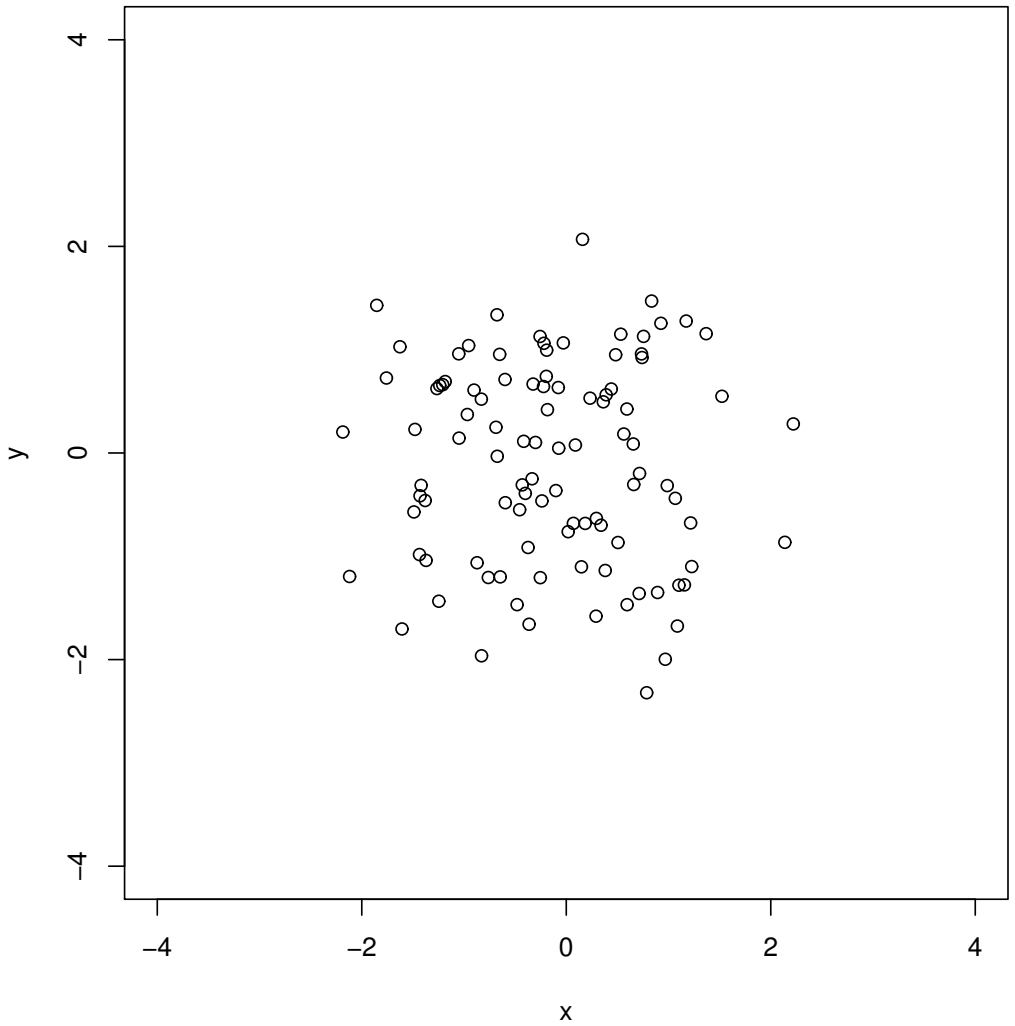
14. Example #1: Good Candidate for PCA



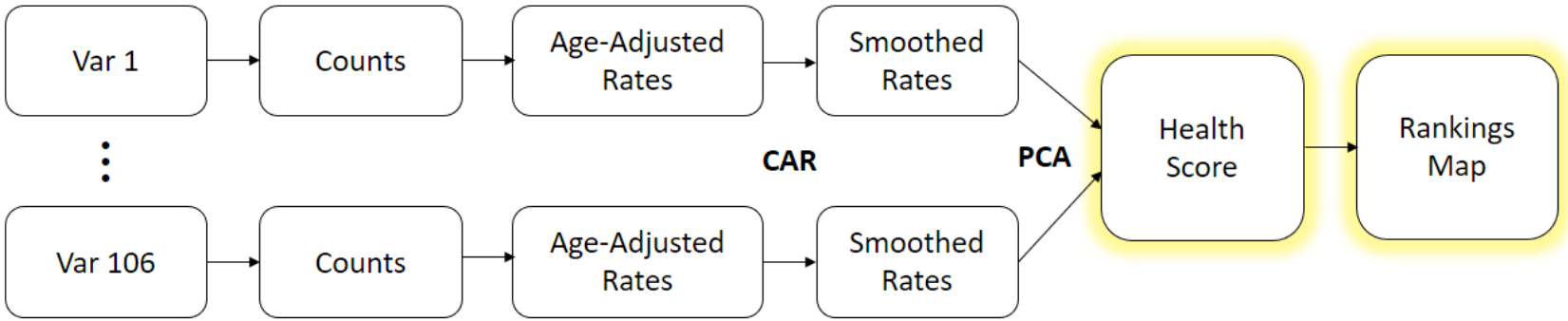
15. Example #1: Good Candidate for PCA



16. Example #2: Bad Candidate for PCA



17. Outline of Code: PCA Health Scores



5. Make function to return health scores using principal components analysis (PCA)
   a. Conduct PCA with the specified health measures
   b. Calculate health score using the first principal component from that PCA
   c. Output a data frame containing each county, their health score, and their ranking

6. Make Shiny application
   - **RUN code**
   - **Discussion**

18. Conclusion

- We performed Markov Chain Monte Carlo (MCMC) simulation with Bayesian CAR modeling to smooth the county health values. Second, we used PCA with the smoothed data to obtain a single health score for each county. Finally, using the health scores, we created an interactive mapping application in order to illustrate which counties are more healthy according to different combinations of health variables.
- This method is useful for creating an overall health score for each county based on selected health measures. It can be used to provide information to counties to help them prioritize the allocation of their resources.
- From the analysis we can conclude that the southern counties have poorer health in general.
- During the analysis, we noticed an issue with the small counts.

19. Future Works

- Incorporate other health measures.
- Examine different years of the Iowa Health Fact Book and analyze the changes in overall health over time.
- Apply this analysis accounting for demographic variability.

20. Acknowledgment

- Dr. Jake Oleson, for mentoring us in the research project.
- Graduate students Marie Ozanne and Jake Clark, for their assistance in learning about PCA and mapping in R.
- Monica Ahrens, for her R coding assistance in the lab sessions.
- Dr. Gideon Zamba, for creating the ISIB program and teaching the lectures.
- ISIB staff and Biostatistics Department, for teaching us more of what it means to be a biostatistician.

21. Special Acknowledgment

- Thanks to the National Heart, Lung, and Blood Institute (NHLBI), for funding ISIB program. (grant #HL131467).

22. References

1. Ringner, M. *What is principal component analysis?* Nature Biotechnology 26, 303 - 304 (2008).
2. Isbell and Littman. *Introduction to Principal Component Analysis (PCA).* (2014).
3. Jolliffe, I.T. *Principal Component Analysis* (2002).
4. Sturtz S, Ligges U, Gelman A. *R2OpenBUGS: a package for running OpenBUGS from R.* (2010).
5. Lunn D, Spiegelhalter D, Thomas A, Best N. *The BUGS project: evolution, critique and future directions.* Stat Med 28:30493067 (2009).