

# Session 1

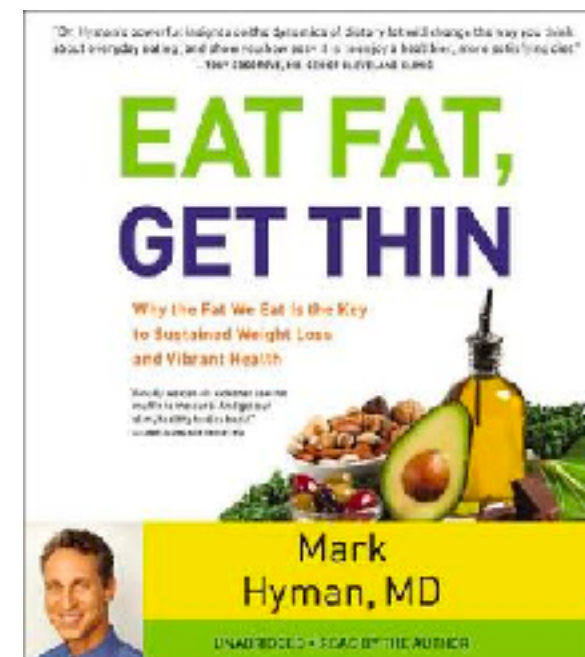
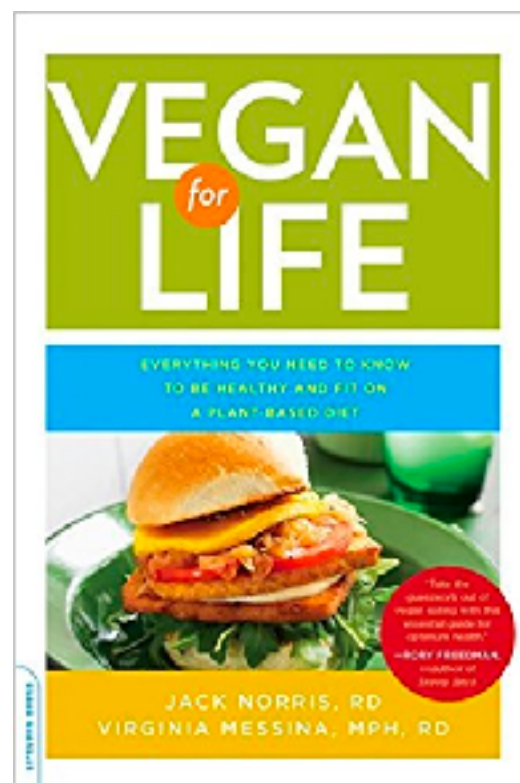
---

Psych 10/Stats 60  
Ismael Lemhadri  
Summer, 2020

Slides available on the website [[stats60.stanford.edu](https://stats60.stanford.edu)]

How do you decide what's healthy to eat?

<https://health.gov/dietaryguidelines/>



## HEALTHY EATING PLATE

**HEALTHY OILS**

Use healthy oils (like olive and canola oil) for cooking, on salad, and at the table. Limit butter. Avoid trans fat.

The more veggies – and the greater the variety – the better. Potatoes and French fries don't count.

Eat plenty of fruits of all colors.

**WATER**

Drink water, tea, or coffee (with little or no sugar). Limit milk/dairy (1-2 servings/day) and juice (1 small glass/day). Avoid sugary drinks.

Eat a variety of whole grains (like whole-wheat bread, whole-grain pasta, and brown rice). Limit refined grains (like white rice and white bread).

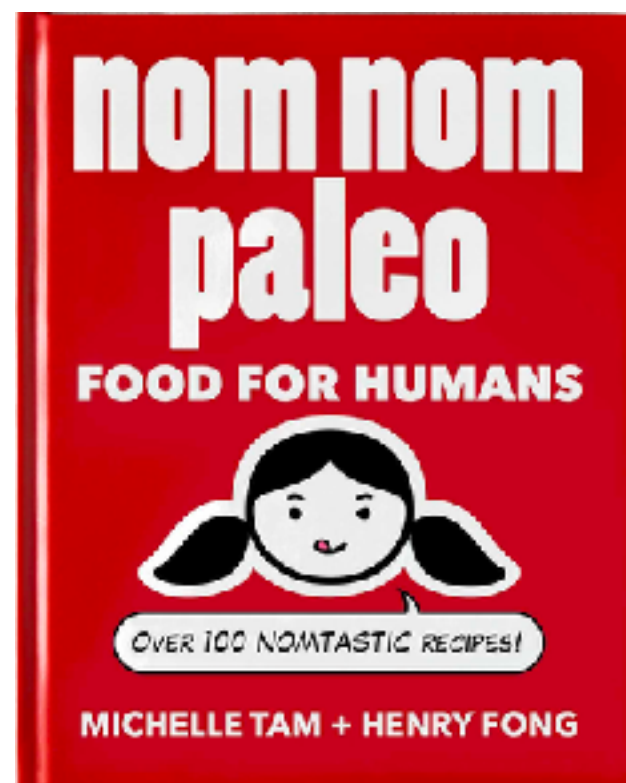
Choose fish, poultry, beans, and nuts; limit red meat and cheese; avoid bacon, cold cuts, and other processed meats.

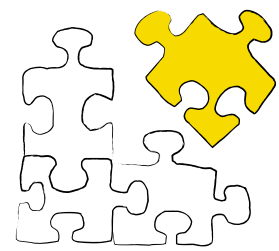
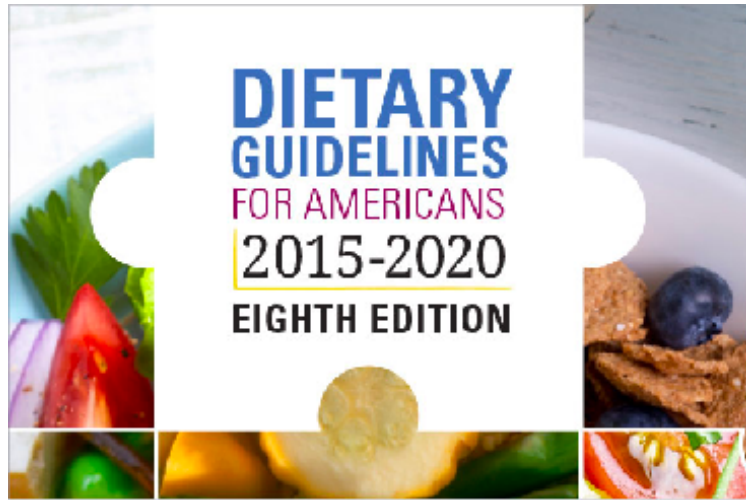
**STAY ACTIVE!**

© Harvard University

Harvard School of Public Health  
The Nutrition Source  
[www.hsph.harvard.edu/nutritionsource](http://www.hsph.harvard.edu/nutritionsource)

Harvard Medical School  
Harvard Health Publications  
[www.health.harvard.edu](http://www.health.harvard.edu)





## Key Recommendations:

**Consume a healthy eating pattern that accounts for all foods and beverages within an appropriate calorie level.**

**A healthy eating pattern includes:<sup>[1]</sup>**

- A variety of vegetables from all of the subgroups—dark green, red and orange, legumes (beans and peas), starchy, and other
- Fruits, especially whole fruits
- Grains, at least half of which are whole grains
- Fat-free or low-fat dairy, including milk, yogurt, cheese, and/or fortified soy beverages
- A variety of protein foods, including seafood, lean meats and poultry, eggs, legumes (beans and peas), and nuts, seeds, and soy products
- Oils

**A healthy eating pattern limits:**

- Saturated fats and *trans* fats, added sugars, and sodium




Why do you think eating saturated fat is unhealthy?



# Common sense vs data

- Common sense
  - Fat in our diet must turn straight into fat in our body, right?
  - Fat clogs pipes, so it must clog our arteries too, right?
- What do the data tell us?

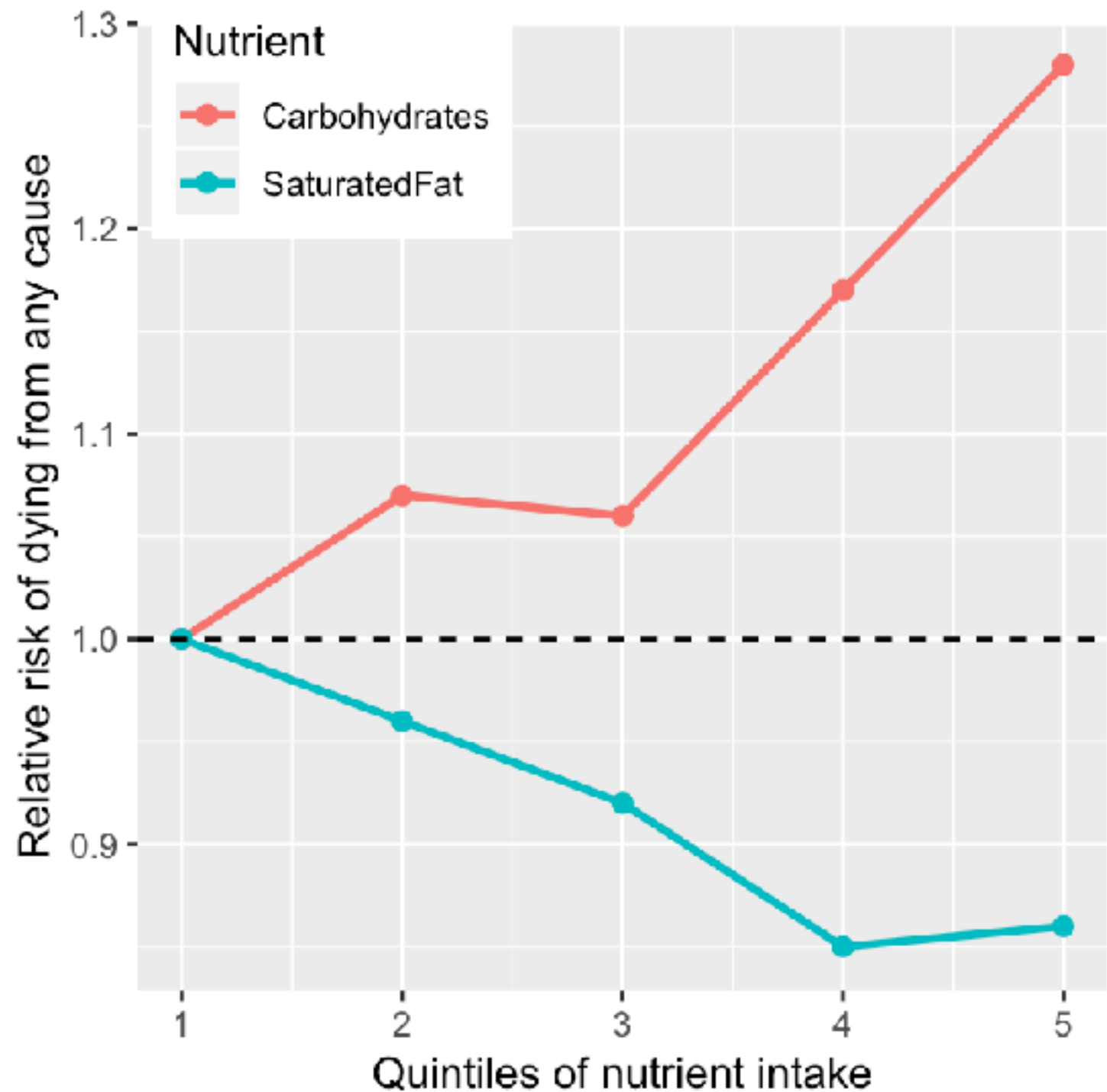


<b>What do your arteries look like?</b>		
		
<b>Healthy artery</b>	<b>Partially blocked artery</b>	<b>Blocked artery</b>
<ul style="list-style-type: none"> <li>– Blood flow</li> <li>– Low risk of heart disease</li> </ul>	<ul style="list-style-type: none"> <li>– Artery narrowed by atherosclerosis</li> <li>– Increased health risk</li> </ul>	<ul style="list-style-type: none"> <li>– Plaque build-up in the lining of artery</li> <li>– Blocked arteries can lead to stroke and heart disease</li> </ul>

PURE study  
Deghan et al, 2017  
The Lancet

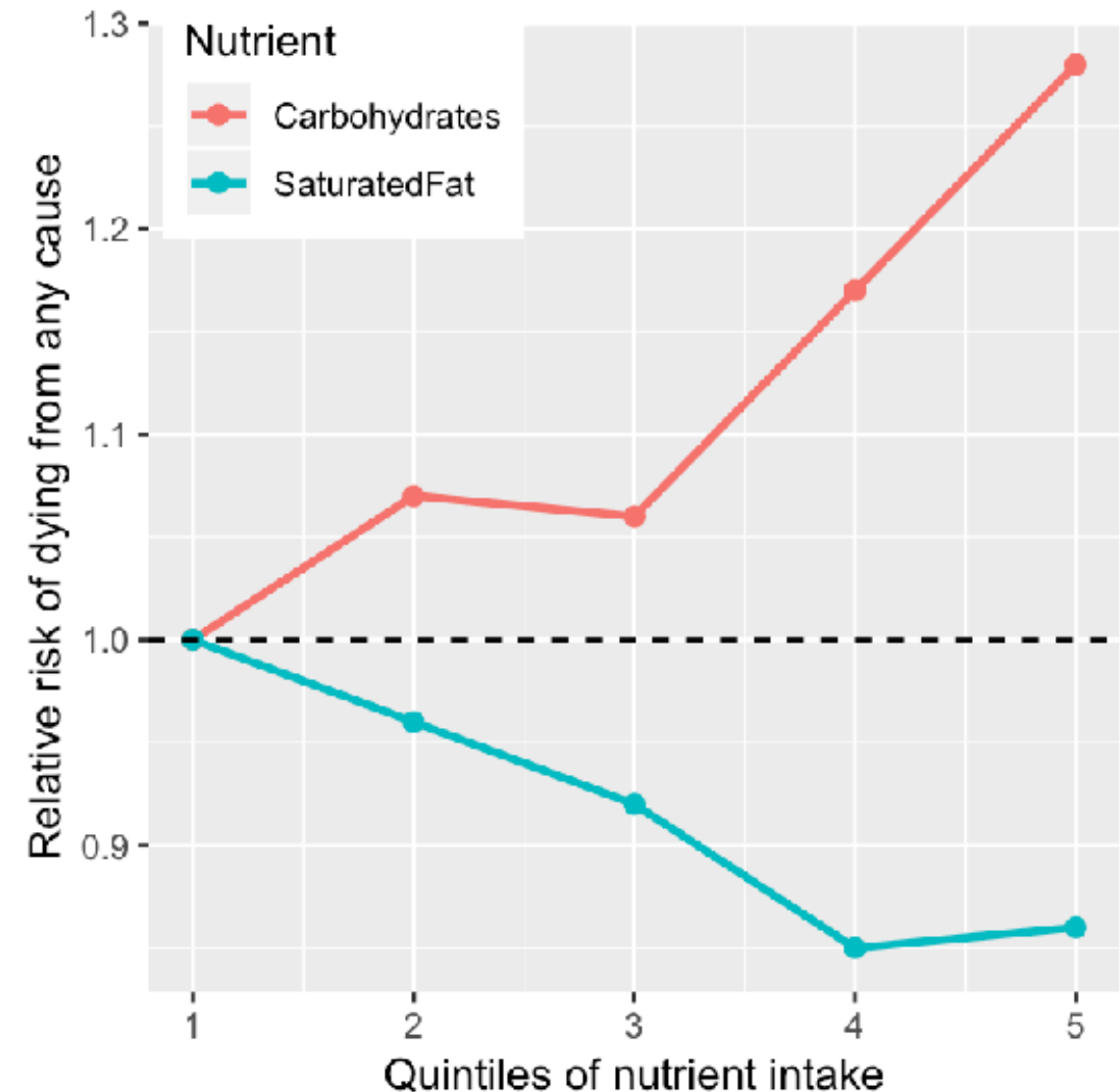
135,335 people  
from 18 countries

followed for median  
7.4 years



# What can statistics do for us?

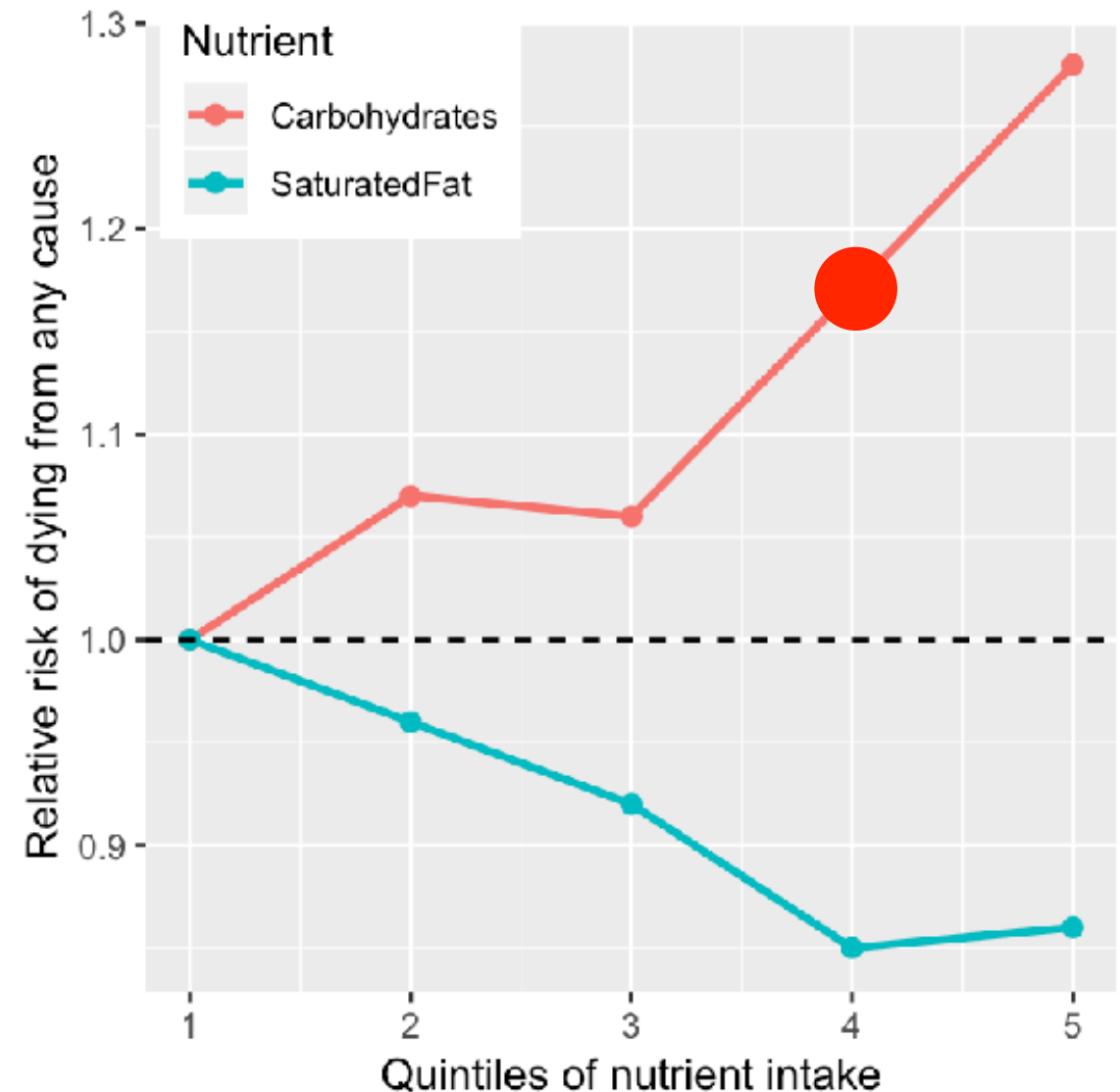
- Describe
  - Provide a summary of a dataset
  - Imagine trying to understand the PURE study by looking at all 135,335 data points!





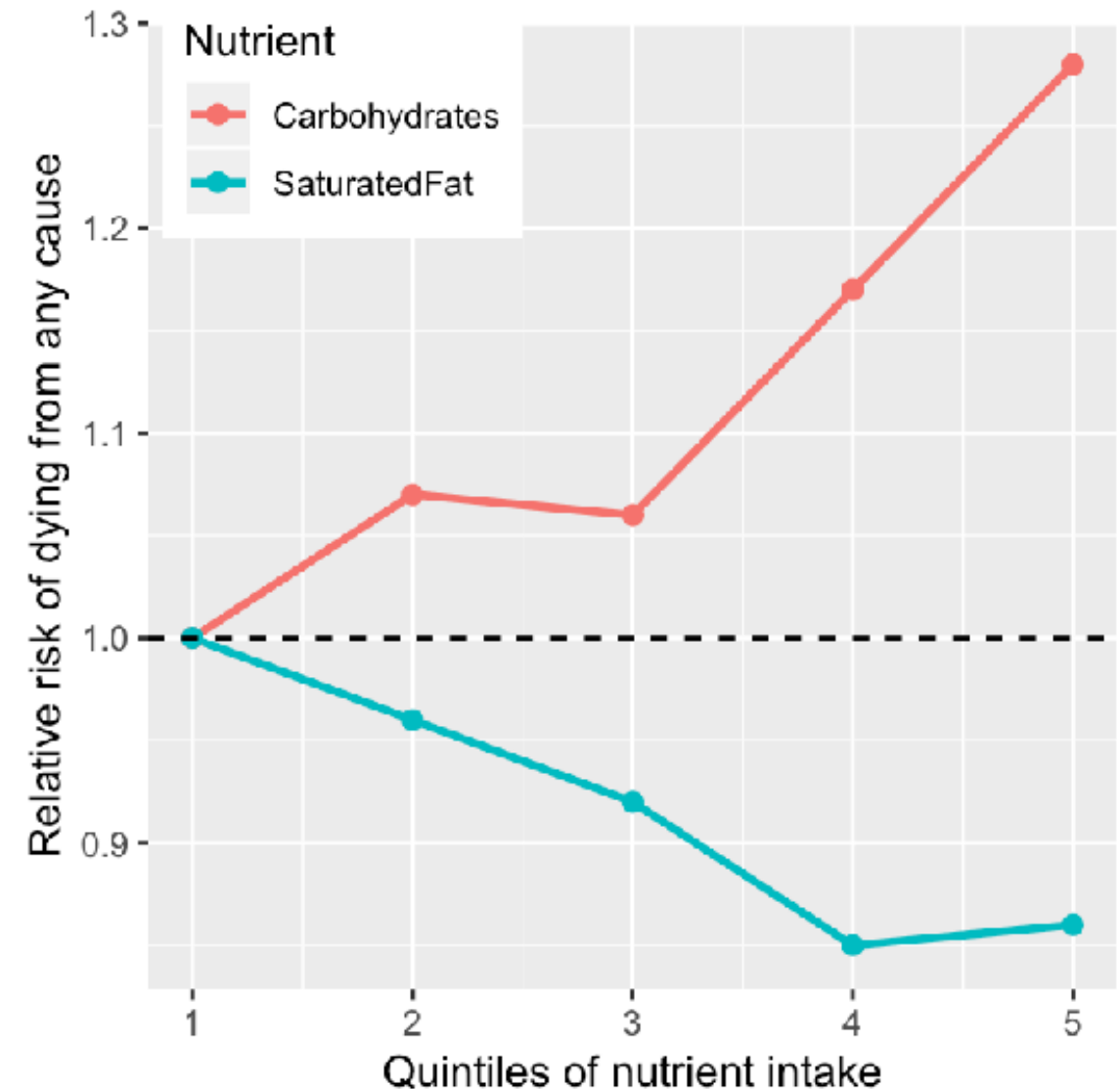
# What can statistics do for us?

- Predict
  - Given some information, predict the most likely outcome
    - Sometimes absolute
      - Predict how many years you will live
    - Sometimes relative
      - Predict your likelihood of dying compared to some reference

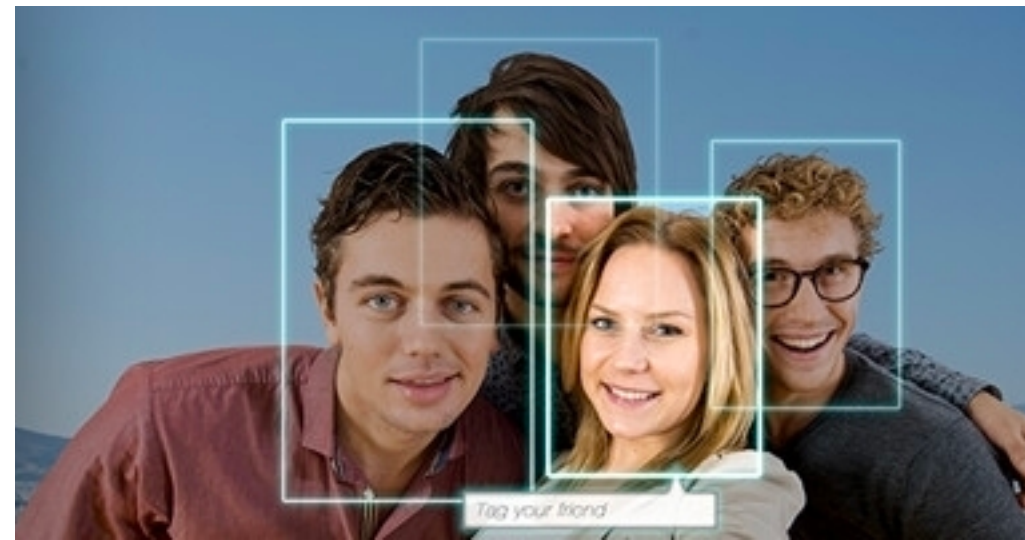


# What can statistics do for us?

- Decide
  - Is there a relationship between fat intake and health?
  - Should the government tell people to eat less or more fat?
- Must consider potential costs and benefits



# Statistics in the real world



## Frequently bought together

Total price: **\$238.09**

[Add all three to Cart](#)

[Add all three to List](#)

Some of these items ship sooner than the others. [Show details](#)

- ✓ **This Item:** Handbook of Functional MRI Data Analysis by Russell A. Poldrack Hardcover **\$74.48**
- ✓ **Statistical Analysis of fMRI Data (MIT Press)** by F. Gregory Ashby Hardcover **\$51.14**
- ✓ **Functional Magnetic Resonance Imaging** by Scott A. Huettel Hardcover **\$112.47**



“Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.”

**- H. G. Wells**

# Data Literacy

---

- The ability to collect, manage, and analyze data is critical to nearly every field of study — not just the sciences!



# Example: Who wrote the Federalist Papers?

- Collection of 85 articles and essays by Alexander Hamilton, James Madison, and John Jay promoting the ratification of the United States Constitution.
- Published in 1787 under the pseudonym "Publius"
- 12 of the papers were later claimed by Madison, but Hamilton's supporters claimed he had written them

In the Press,  
and speedily will be published,  
**THE**  
**FEDERALIST,**  
A Collection of Essays written in fa-  
vor of the New Constitution.  
By a Citizen of New-York.  
Corrected by the Author, with Additions  
and Alterations.

*This work will be printed on a fine Paper  
and good Type, in one handsome Volume duo-  
decimo, and delivered to subscribers at the  
moderate price of one dollar. A few copies  
will be printed on superfine royal writing pa-  
per, price ten shillings.*

*No money required till delivery.*

*To render this work more complete, will be  
added, without any additional expence,*

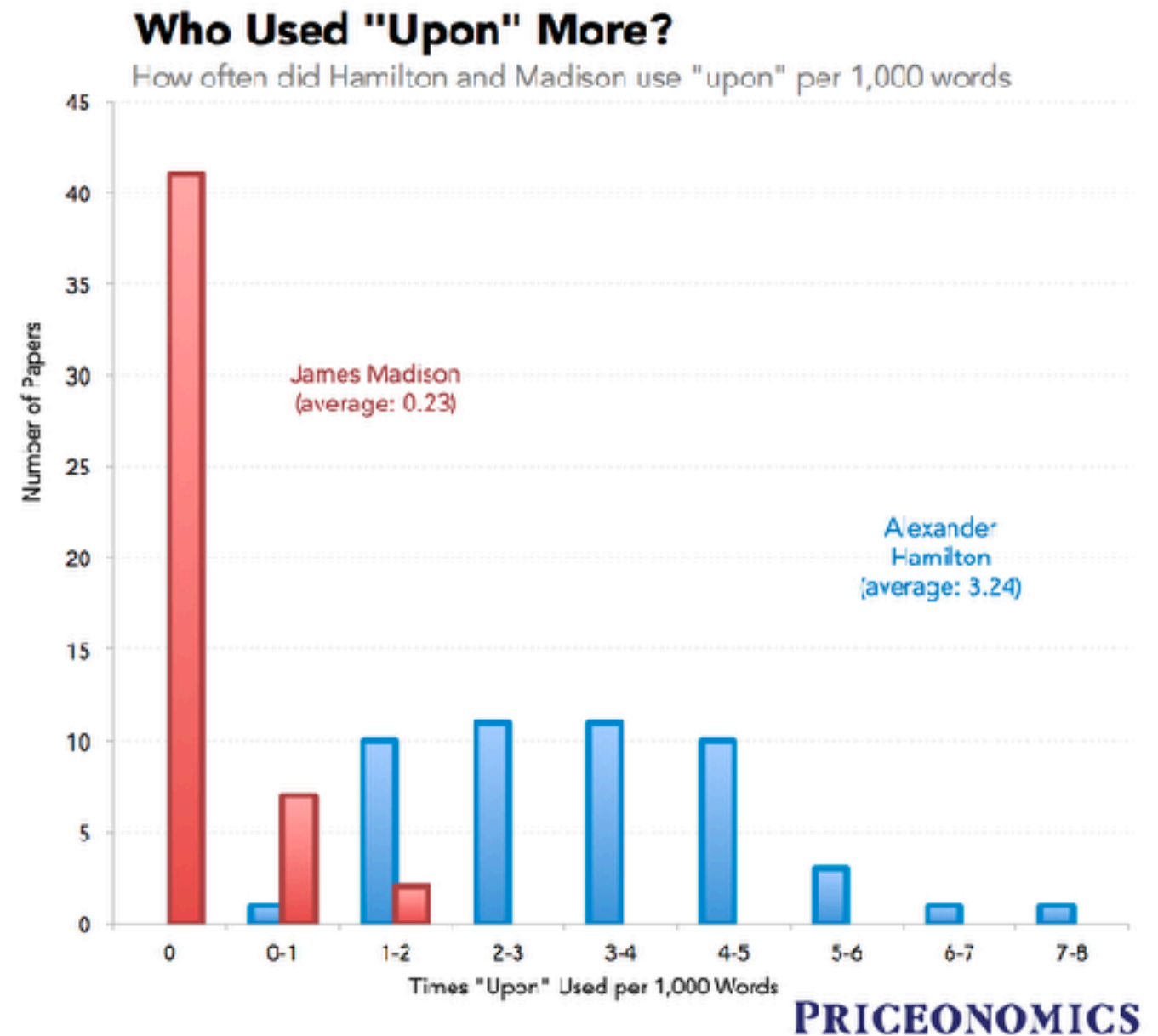
**PHILO-PUBLIUS,**  
**AND THE**  
*Articles of the Convention,*  
*As agreed upon at Philadelphia, Septem-  
ber 17th, 1787.*

SHEPHERD

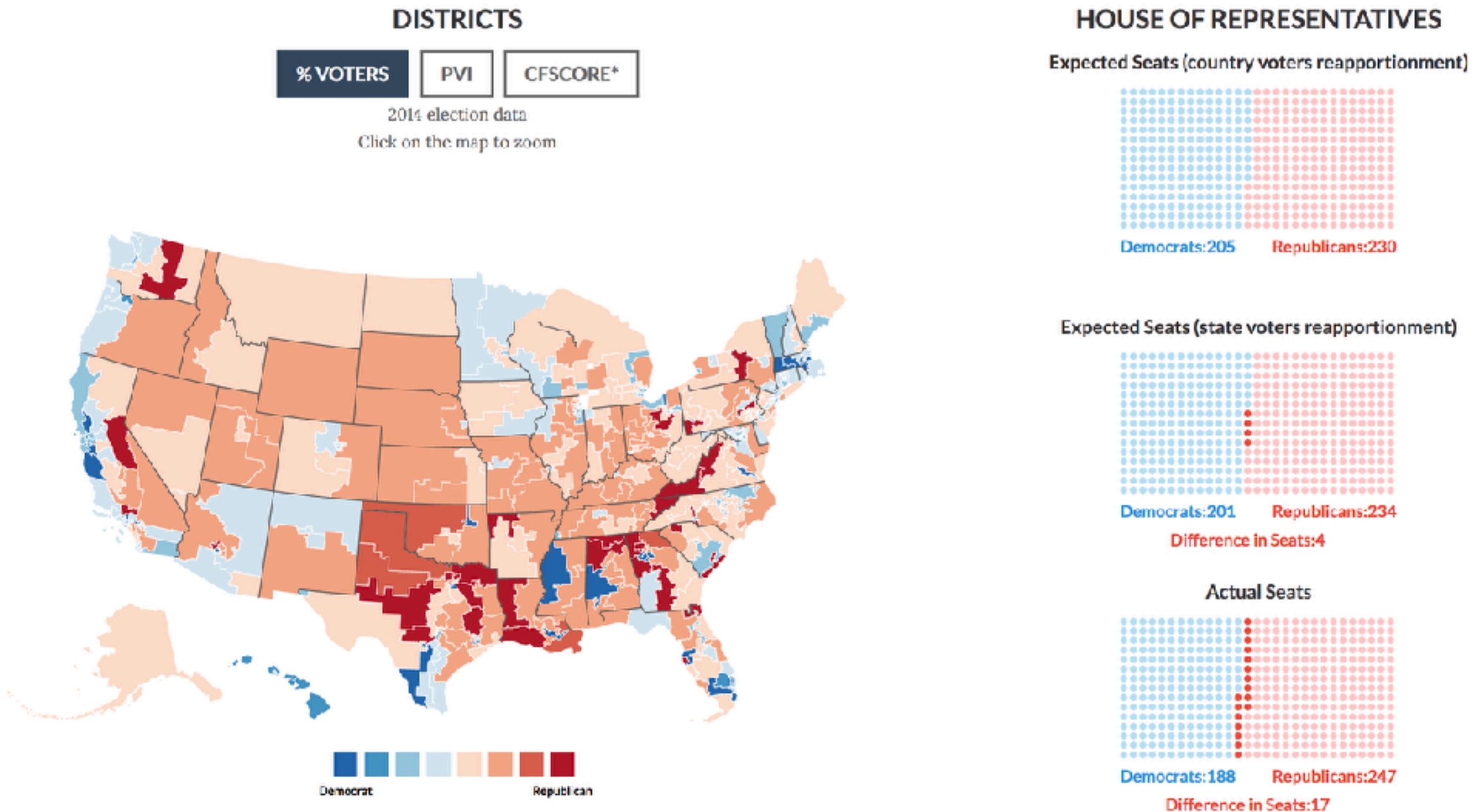


# Example: Who wrote the Federalist Papers?

- In 1959, Mosteller & Wallace analyzed the use of words in the texts, and showed that the odds were 12.6 million to 1 that Madison had written the 12 disputed papers

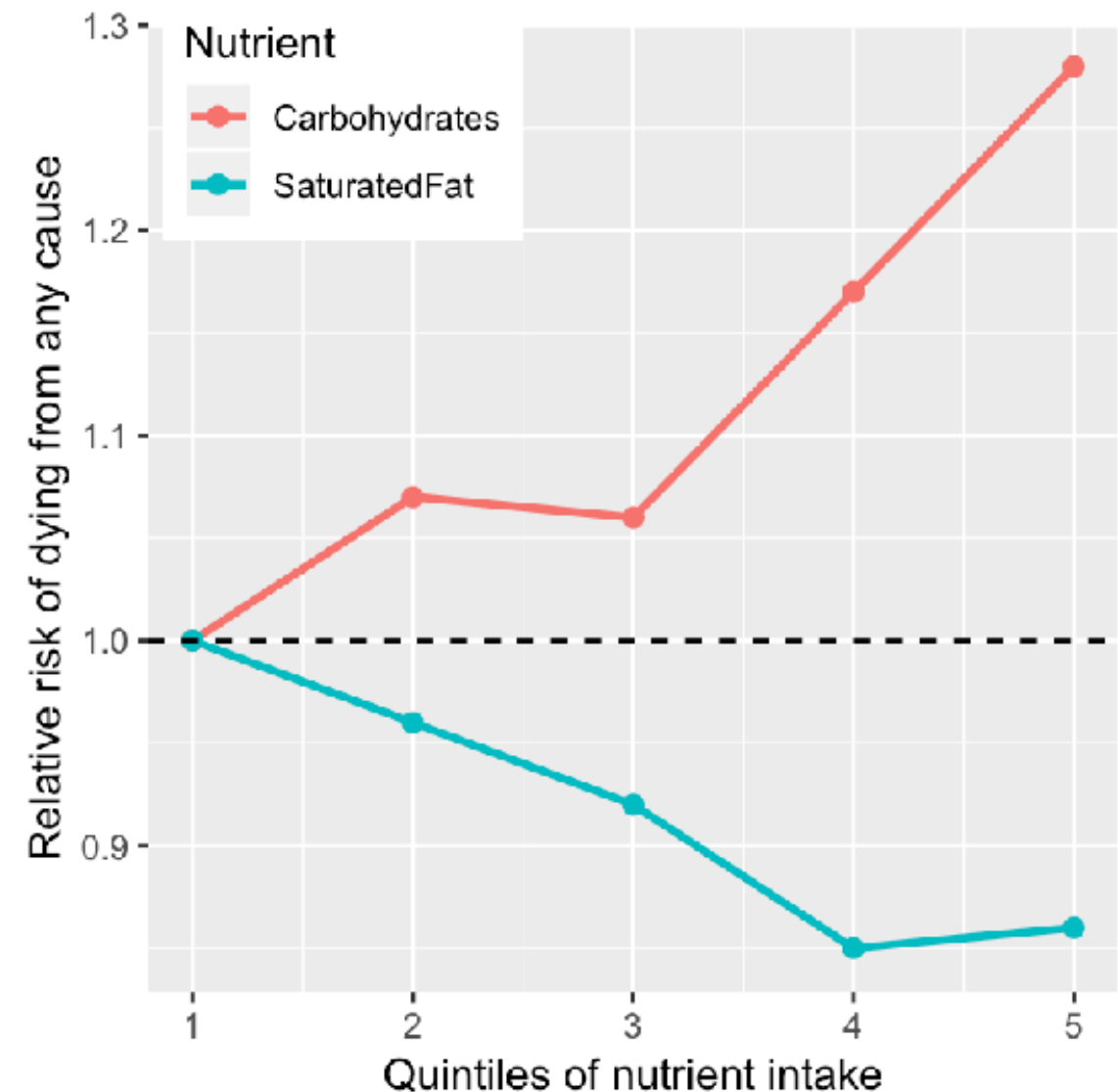


# Example: Quantifying effects of gerrymandering



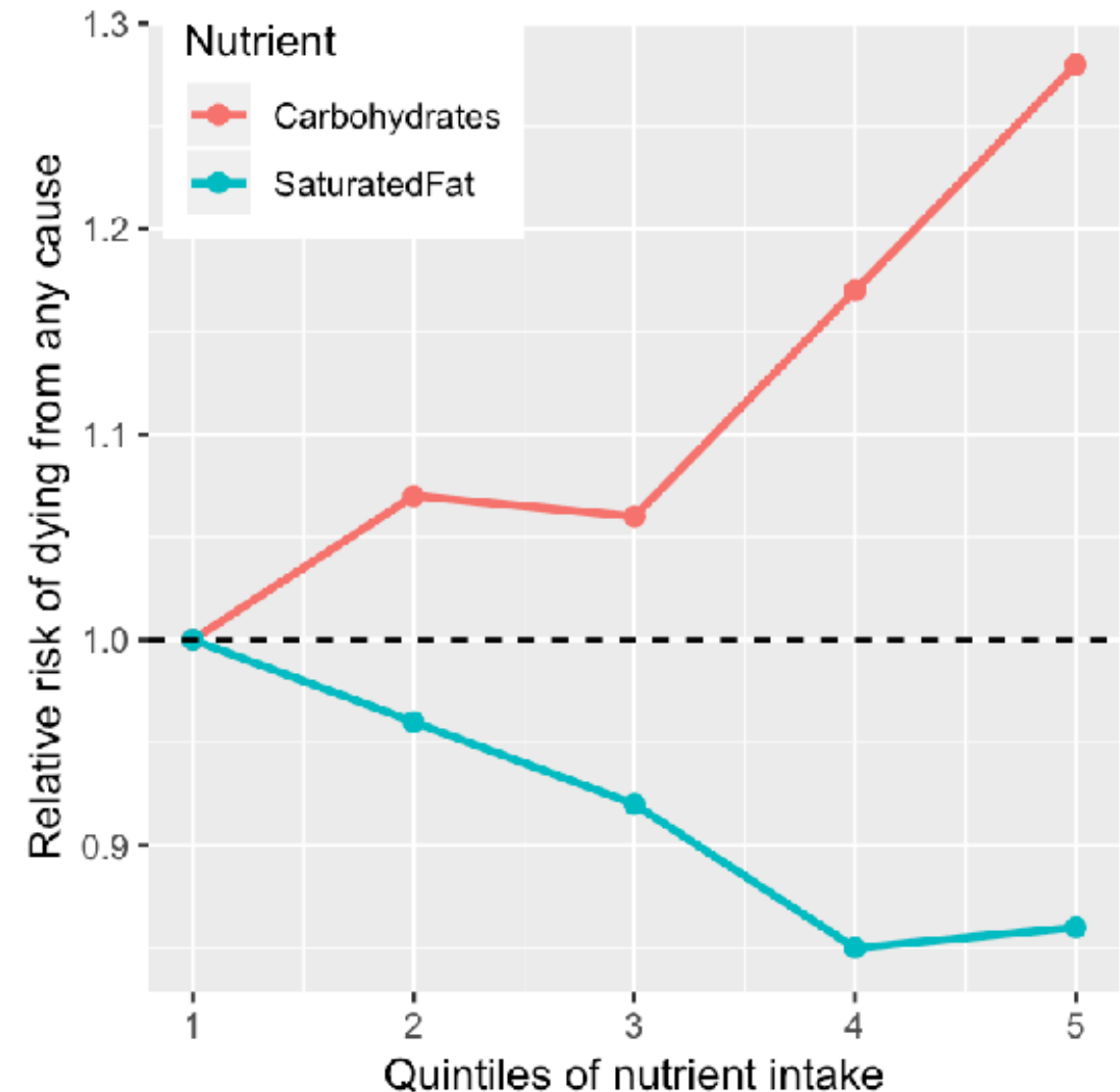
# Fundamental concepts of statistics

- Learning from data
  - We come to any question with prior beliefs
  - We update or change those beliefs based on the data
- “Machine learning”



# Fundamental concepts of statistics

- Aggregation
  - We can actually gain understanding by throwing away information
  - We don't need to know most of the details about each of the 135,535 people
  - We just need the aggregated information



# Fundamental concepts of statistics

- Uncertainty
  - The world is probabilistic
    - Smoking causes lung cancer, but some smokers will never get the disease
  - We can't be sure, but we can quantify our uncertainty
    - a 68-year-old man who smoked two packs a day for the past 50 years and continues to smoke has a 15% (1 out of 7) risk of getting lung cancer

NEWS

## This 112-year-old woman smokes 30 cigarettes a day

By David K. Li

January 26, 2016 | 12:04pm

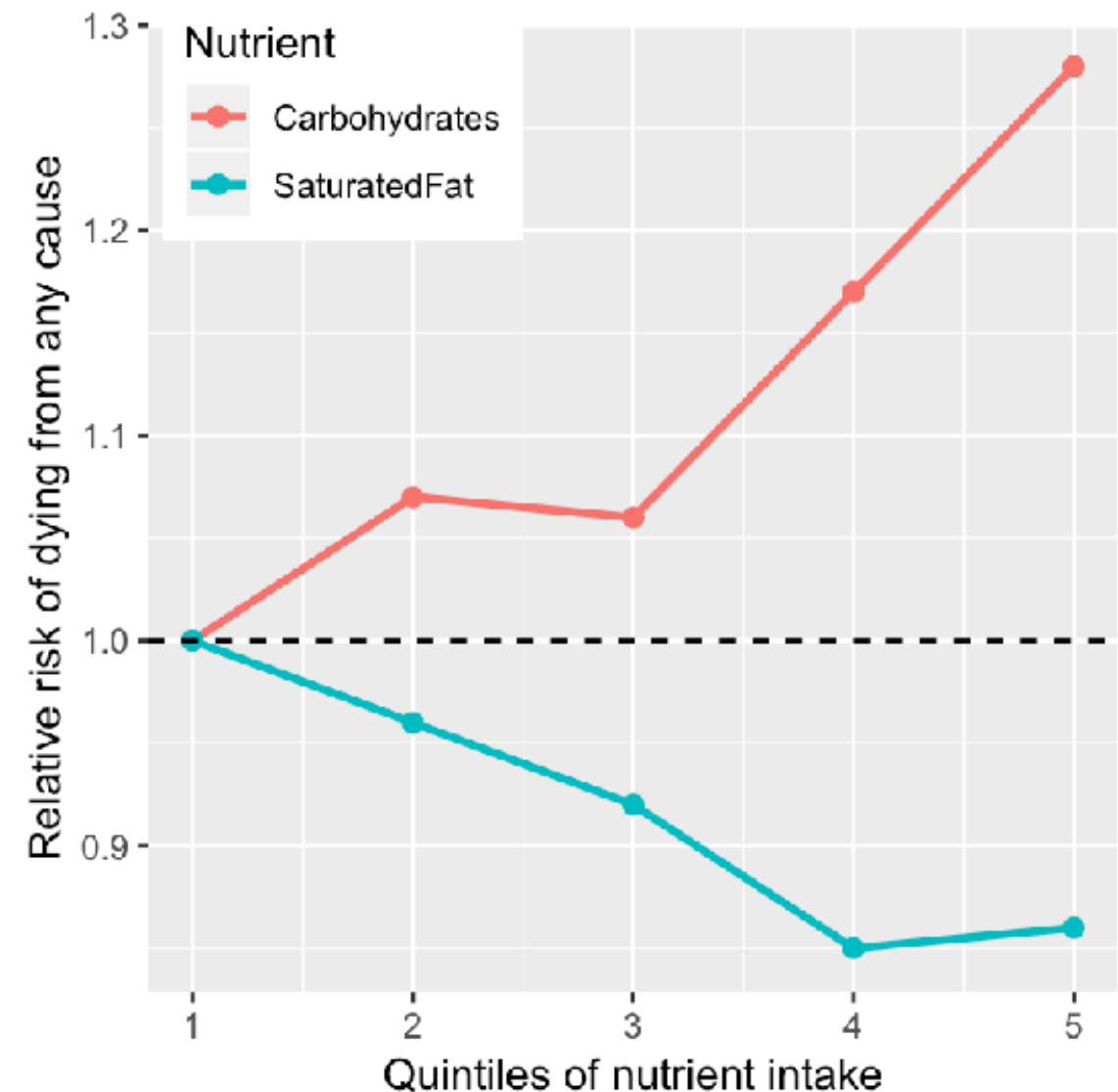


Batuli Lamichhane has credited her longevity to chain smoking.



# Fundamental concepts of statistics

- Sampling
  - We can make generalizations about 7.5 million humans in 195 countries based on a sample of ~135K people from 18 countries
  - Bigger samples help, but there are diminishing returns as they get bigger
  - The sample has to be appropriate
    - Sampling only from countries with a high proportion of vegetarians could bias the results

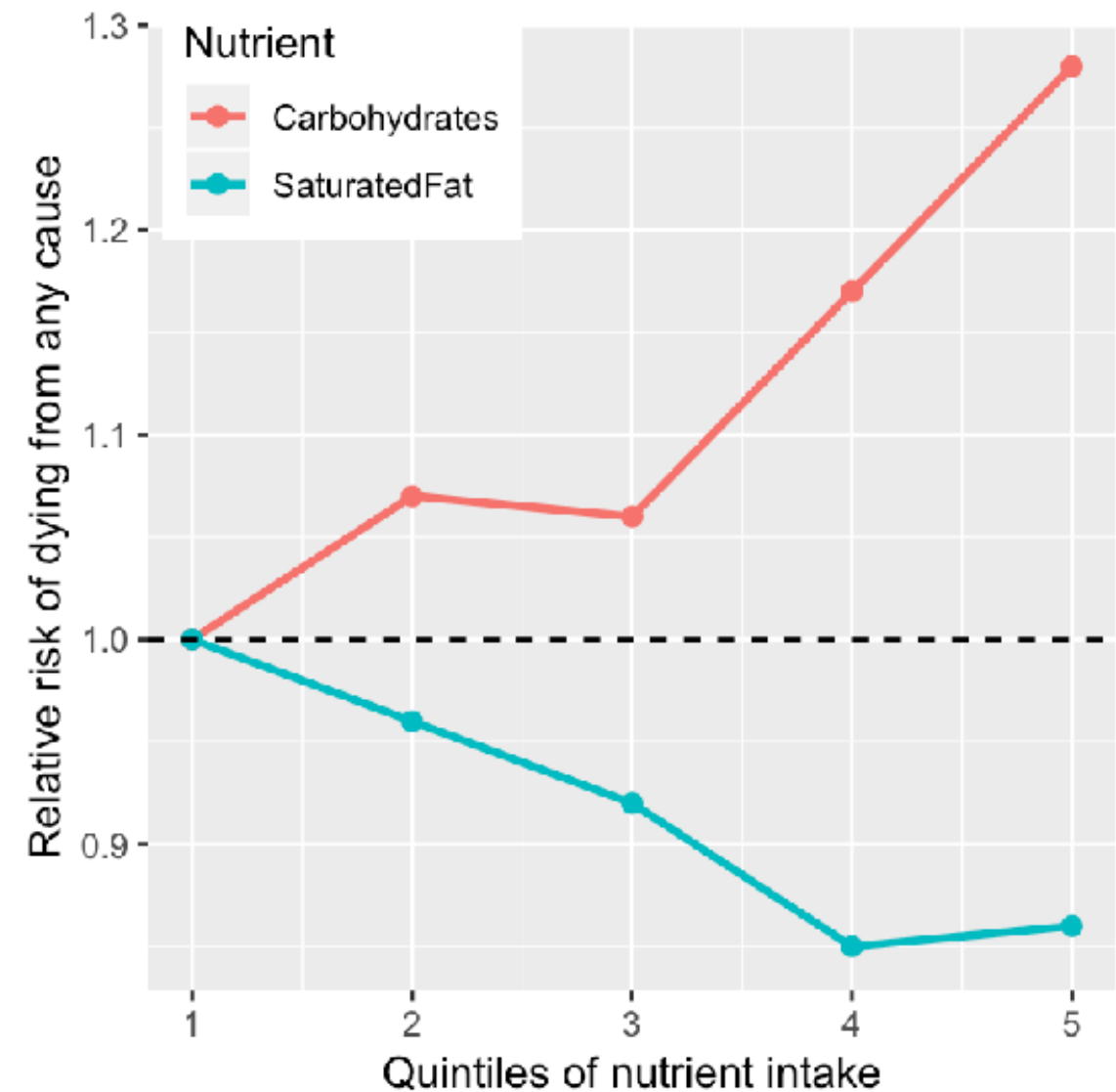




Does this result tell us that eating fat *causes* reduced mortality and disease?

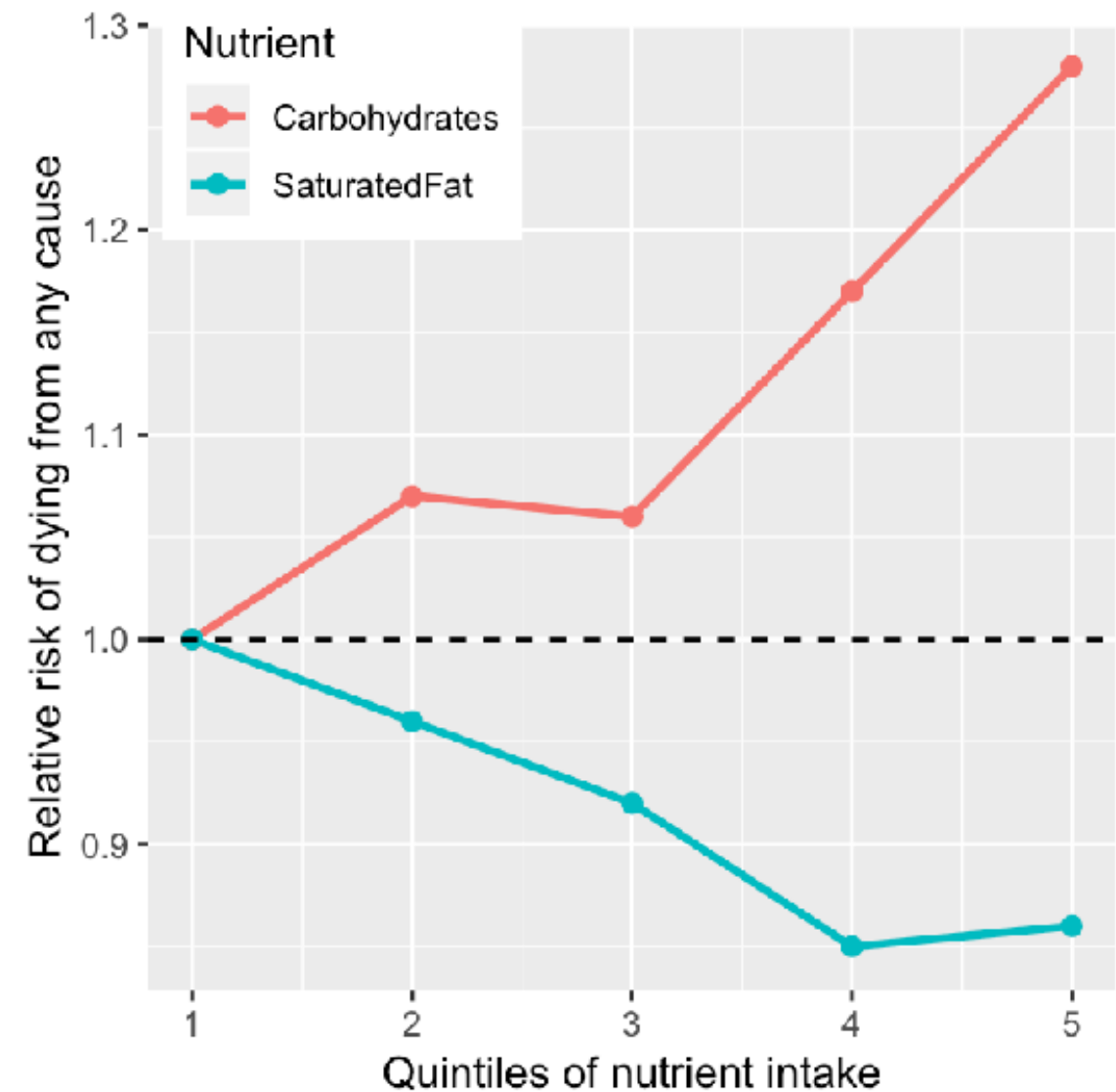
If so, how?

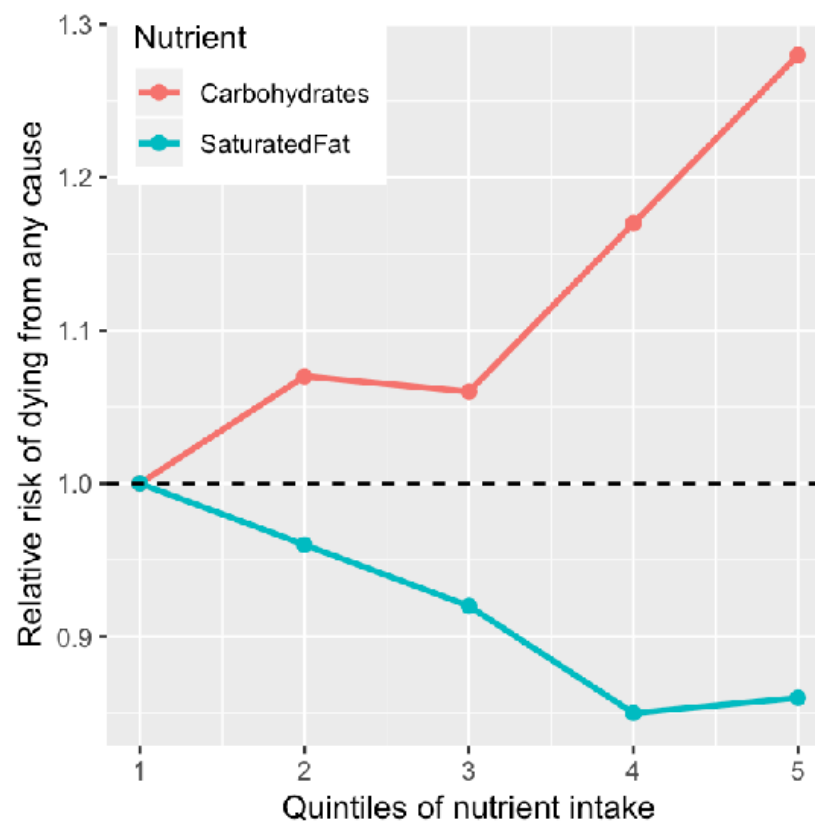
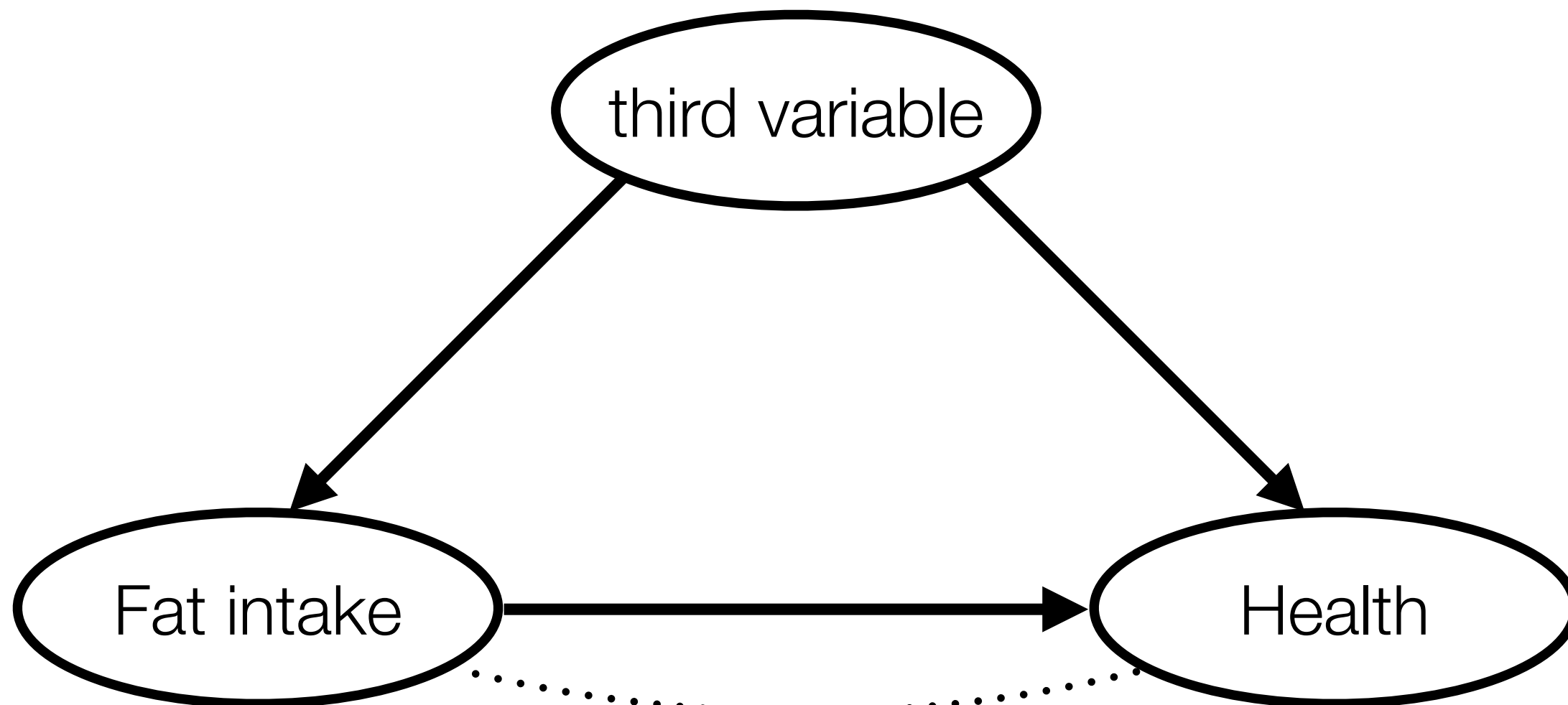
If not, why not?



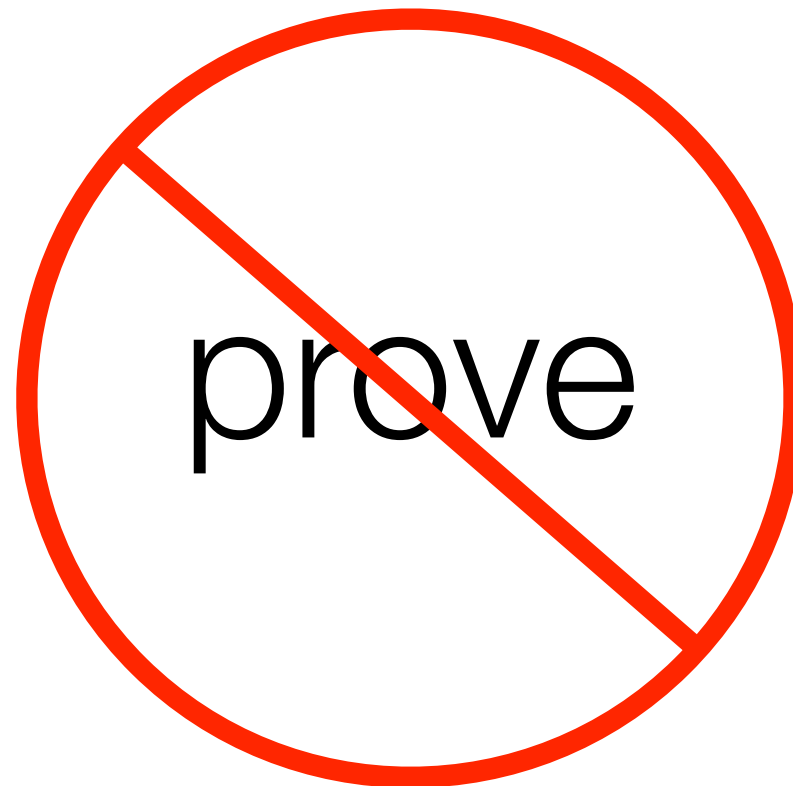
**Figure 1: Association between estimated percentage energy from nutrients and total mortality and major cardiovascular disease (n=135 335)**

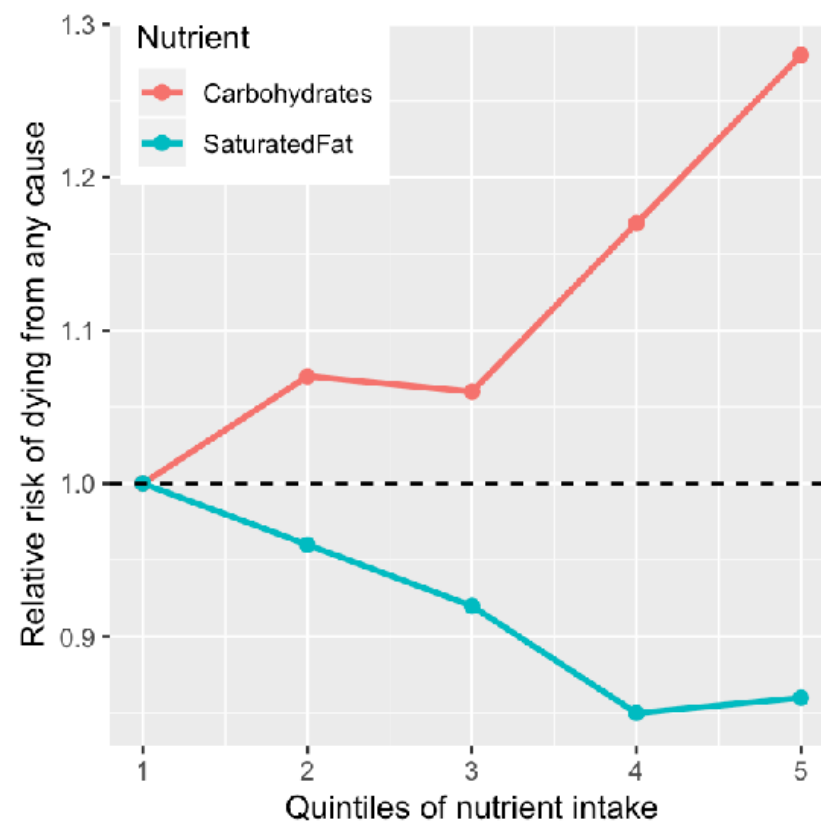
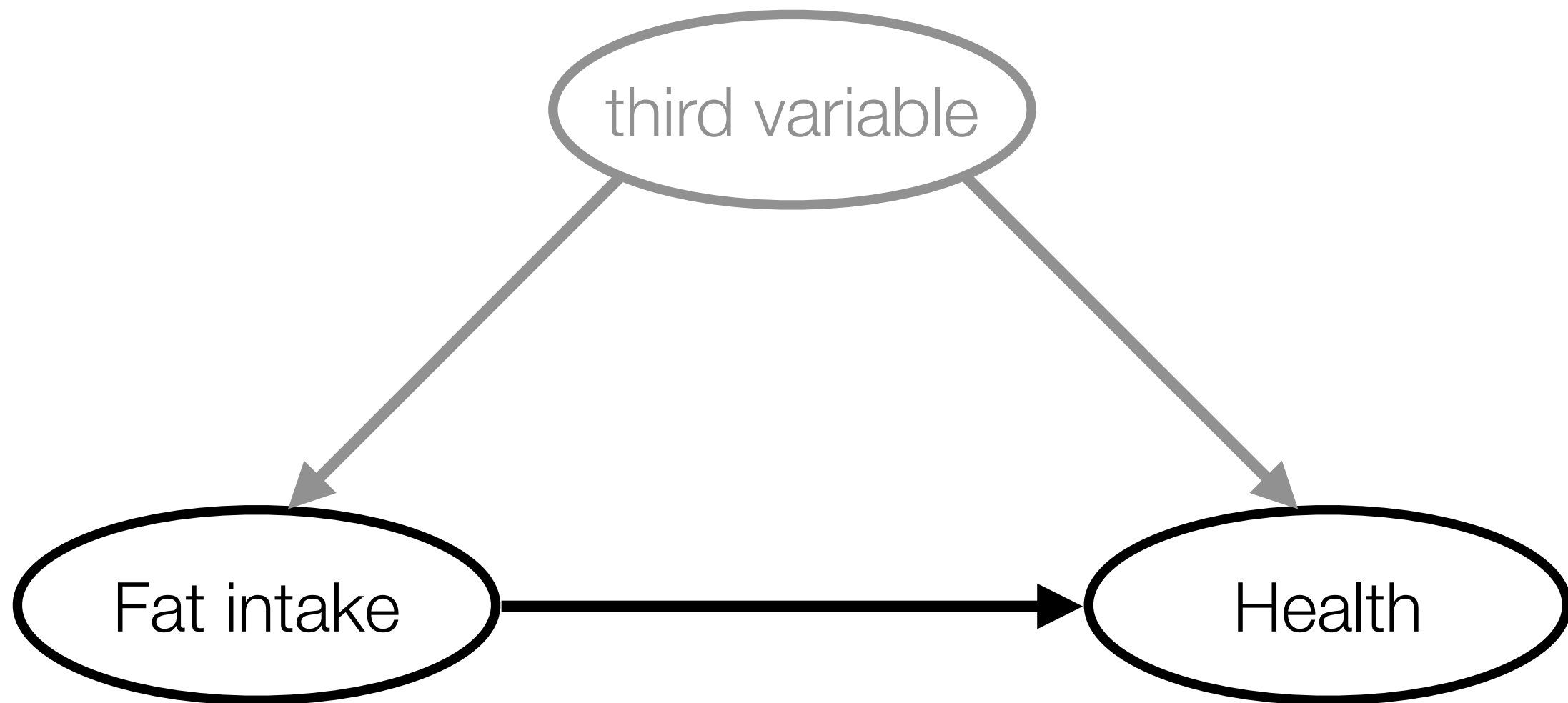
Adjusted for age, sex, education, waist-to-hip ratio, smoking, physical activity, diabetes, urban or rural location, centre, geographical regions, and energy intake.





How could we establish with more certainty that eating fat causes reduced disease and death?

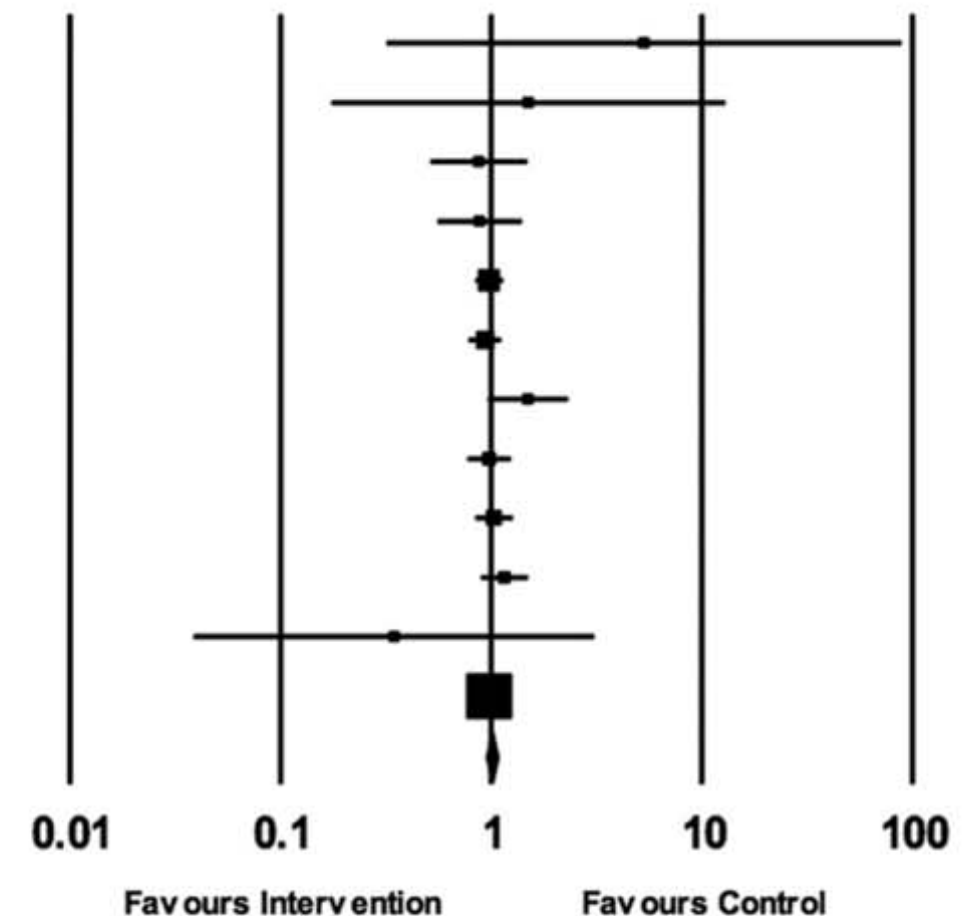




# Randomized controlled trials of dietary fat reduction

## Dietary Intervention & All Deaths

Study name	Risk			All Deaths / Total		Risk ratio and 95% CI
	ratio	Lower limit	Upper limit	Intervention	Control	
Rose Corn Oil (1965)	5.310	0.315	89.441	5/28	0/13	
Rose Olive Oil (1965)	1.500	0.172	13.046	3/26	1/13	
Research Committee Low-Fat (1965)	0.874	0.510	1.499	20/123	24/129	
MRC Soybean Oil (1968)	0.881	0.550	1.411	28/199	31/194	
LA Veterans Dayton (1969)	0.978	0.834	1.148	174/424	177/422	
Leren, Oslo heart study (1970)	0.935	0.773	1.131	101/206	108/206	
Woodhill, Sydney heart study (1978)	1.494	0.953	2.342	39/221	28/237	
DART Fat advice (1989)	0.979	0.765	1.254	111/1018	113/1015	
Frantz Minnesota Men (1989)	1.032	0.833	1.279	158/2197	153/2196	
Frantz Minnesota Women (1989)	1.156	0.885	1.512	111/2344	95/2320	
STARS (1992)	0.346	0.038	3.122	1/27	3/28	
WHI (2006)	0.979	0.904	1.061	950/19541	1454/29294	
	0.991	0.935	1.051			





Questions?

# Overview of the course: Instructional team

---

- Instructor: Ismael Lemhadri (he/him)
- Teaching assistants
  - Yu Wang (he/him)
  - Yuchen Wu (she/her)
  - Qian Zhao (she/her)

# Overview of the class: Web sites

---

- Primary course web site:
  - <http://stats60.stanford.edu>
- **Canvas** for everything Zoom
- **Piazza**
  - Primary discussion forum for class
  - You must self-enroll
- **Gradescope**

# Overview of the course

**Rationale:** It is impossible to understand the modern world without an understanding of statistics. From public opinion polls to clinical trials in medicine to online systems that recommend purchases to us, statistics play a role in nearly every aspect of our lives. The goal of this course is to provide an understanding of the essential concepts and “big ideas” of statistics — describing data and making decisions and predictions based on data — as well as the skills to employ these concepts on real data to solve authentic problems. At the end of the course, students will possess:

- *Statistical literacy:* The ability to dissect and understand statistical claims in scientific research and popular media
- *Statistical ability:* The skills necessary to apply statistical analysis methods to real data
- *Statistical curiosity:* The interest in further developing their statistical skills and knowledge, and the confidence in their ability to do so

# Communication with TAs and each other

## **Piazza**

- Post comments and questions (searchable)
- Private posts
- If you did not receive an email, sign up with link on the syllabus

## **Email ([lemhadri@stanford.edu](mailto:lemhadri@stanford.edu))**

- Use only for questions specific to you

**Each week a TA will monitor Piazza on a daily basis.**

# Overview of the class: Code of conduct

---

- You are expected to treat the instructional team and your fellow students with courtesy and respect.
- This class should be a harassment-free learning experience for everyone regardless of gender, gender identity and expression, sexual orientation, disability, physical appearance, body size, race, age or religion. Harassment of any form will not be tolerated.
- If someone makes you or anyone else feel unsafe or unwelcome, please report it as soon as possible to one of the instructors.
- You may also speak to the Stanford Office of the Ombuds (<https://ombuds.stanford.edu/>)



# Overview of class: Honor Code

---

- Unless otherwise stated, you can use any published resource you wish to complete the assessments (textbook, Internet, etc).
- However, you should not discuss the answers with your fellow students in person or electronically unless instructed to do so by the instructors; sharing answers (including computer code) will be viewed as a violation of the Honor Code.
- Problem sets will be checked automatically for duplicated answers. Last year this method caught a number of students who were sharing answers with one another
- Anyone caught violating these rules will be reported to the Office of Community Standards.

Free open-source textbook

---

## **Statistical Thinking for the 21st Century**

<http://statsthinking21.org/>

All code is available at

<https://github.com/statsthinking21/statsthinking21-core>

If you find errors or have suggestions, you can submit them at  
<https://github.com/statsthinking21/statsthinking21-core/issues>

**0.1** Why does this book exist?

**0.2** You're not a statistician - why sh...

**0.3** Why R?

**0.4** The golden age of data

**0.5** An open source book

**0.6** Acknowledgements

## 1 Introduction

**1.1** What is statistical thinking?

**1.2** What can statistics do for us?

**1.3** Fundamental concepts of statist...

**1.4** Causality and statistics

**1.5** Suggested readings

## 2 Working with data

**2.1** What are data?

**2.2** Scales of measurement

**2.3** What makes a good measureme...

**2.4** Suggested readings

## 3 Probability

**3.1** What is probability?

PDF

EPUB

# Statistical Thinking for the 21st Century

*Copyright 2018 Russell A. Poldrack*

*Draft: 2018-12-11*

## Preface

### 0.1 Why does this book exist?

In 2018 I began teaching an undergraduate statistics course at Stanford (Psych 10/Stats 60). I had never taught statistics before, and this was a chance to shake things up. I have been increasingly unhappy with undergraduate statistics education in psychology, and I wanted to bring a number of new ideas and approaches to the class. In particular, I wanted to bring to bear the approaches that are increasingly used in real statistical practice in the 21st century. As Brad Efron and Trevor Hastie laid out so nicely in their book "Computer Age Statistical Inference: Algorithms, Evidence, and Data Science", these methods take advantage of today's increased computing power to solve statistical problems in ways that go far beyond the more standard methods that are usually taught in the undergraduate statistics course for psychology students.

# What do I expect of you?

---

- Show up to lecture if you can
- Do the work
- Be engaged
- Give us honest feedback

# Hands-on statistics with R

---

- We will use the freely available RStudio statistical software
- Why?

The image shows a snippet of R code with a large, semi-transparent blue 'R' logo overlaid on it. The code includes functions like 'na.omit', 'if', 'else', 'stocks\$close', 'stocks\$open', 'sample', and 'any'.

```
stocks <- na.omit(stocks)
if (stocks$close > stocks$open) {
  status = "up"
} else {
  status = "down"
} else if (stocks$close == stocks$open) {
  status = "flat"
}
stocks$status = status
cor(stocks$close, stocks$open)
table(status)
x <- sample(1:10, 10)
y <- sample(1:10, 10)
all(x > 0)
any(x == y)
```

Learning stats is like learning a sport or musical instrument - the only way to get better is practice!

# Questions?

---

- Spend the next few minutes reading the syllabus at [stats60.stanford.edu](https://stats60.stanford.edu) and discussing with your neighbors
- Write down any questions you have

# Action items

---

- Complete Assignment 0 by Thursday:
  - Enter your stanford.edu email address so that you will get credit!
- Complete reading: Chapters 1 and 2