



Andrey Rzhetsky  
Андрей Ржецкий



# Adventures with large biomedical datasets: environmental exposures and disease

University of Chicago





Portrait of  
Princess  
Adelaide de  
Savoye

# Maria-Adelaide de Savoye

6 December 1685 – 12  
February 1712 (died @ 26!)

Married in 1697 (@ 12  
age at the portrait)

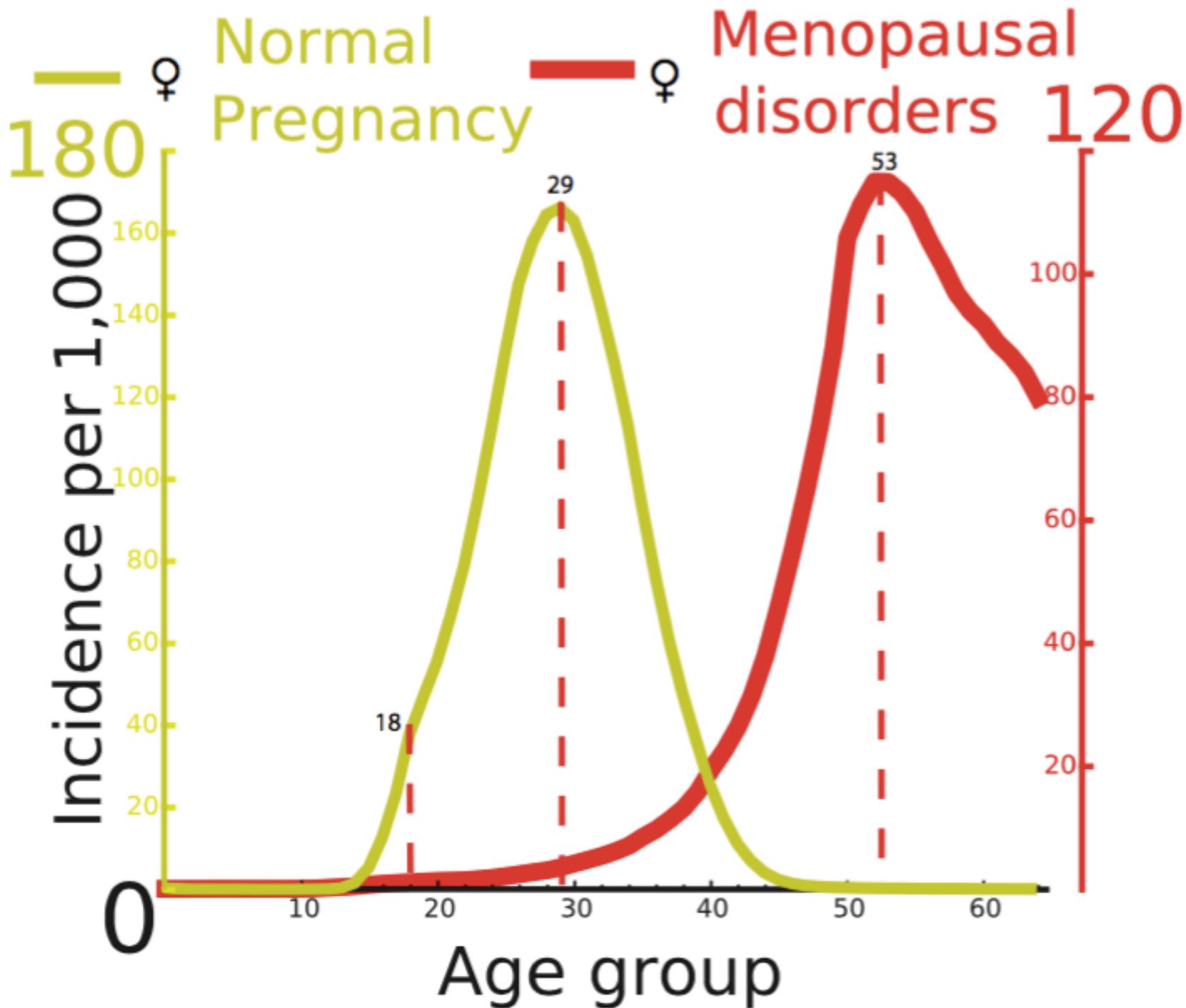


Princess  
Adelaide de  
Savoie

In early April 1711, her father-in-law, Le Grand Dauphin, caught smallpox and died on April 14.

Upon the death of Le Grand Dauphin, Marie Adélaïde's husband became Dauphin of France and she -- Dauphine of France.

At Fontainebleau, Marie Adélaïde caught a fever which escalated in measles -- she died in Versailles at the age of 26. The Dauphin himself died six days later, having caught the measles from his wife.





Data that we used in  
the studies that I will  
cover

International Classification of  
Disease (ICD): versions 8, 9,  
10, 11

# Life of a patient

Current age:

15

Ethnicity:  
Hispanic

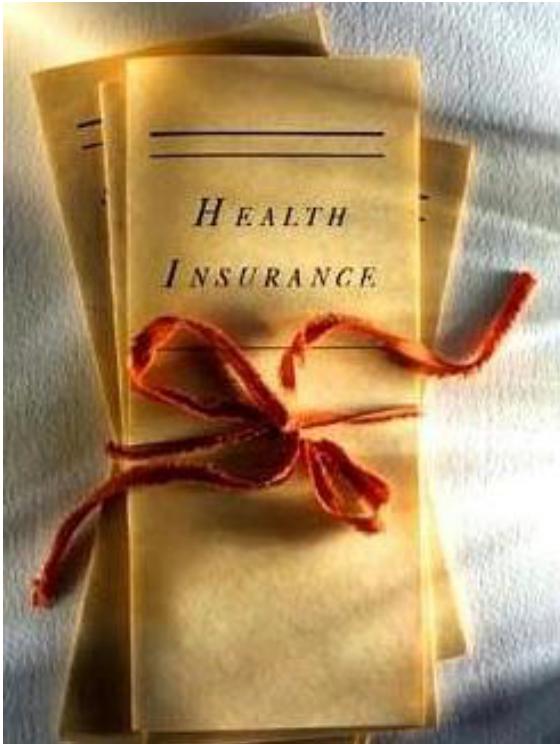
Female

Unspecified viral  
Infection at age 6

15|H|F^079.9:5|079.99:6|278.0:3|345.10:6|372.30:0|  
389.9:4|462:7|465.9:2|474.0:6|474.11:6|478.1:6|486:7|  
493.90:7|784.0:5|785.1:7|786.50:5|959.8:1|999.99:0|  
V20.1:5|V20.2:13|V21.2:4|V30.00:0|V62.89:6|V67.0:4|  
V67.9:0|V70.0:3|V72.1:5

27|U|M^079.9:8|493.90:10 ...

Observed data: ICD9 codes in a clinical database

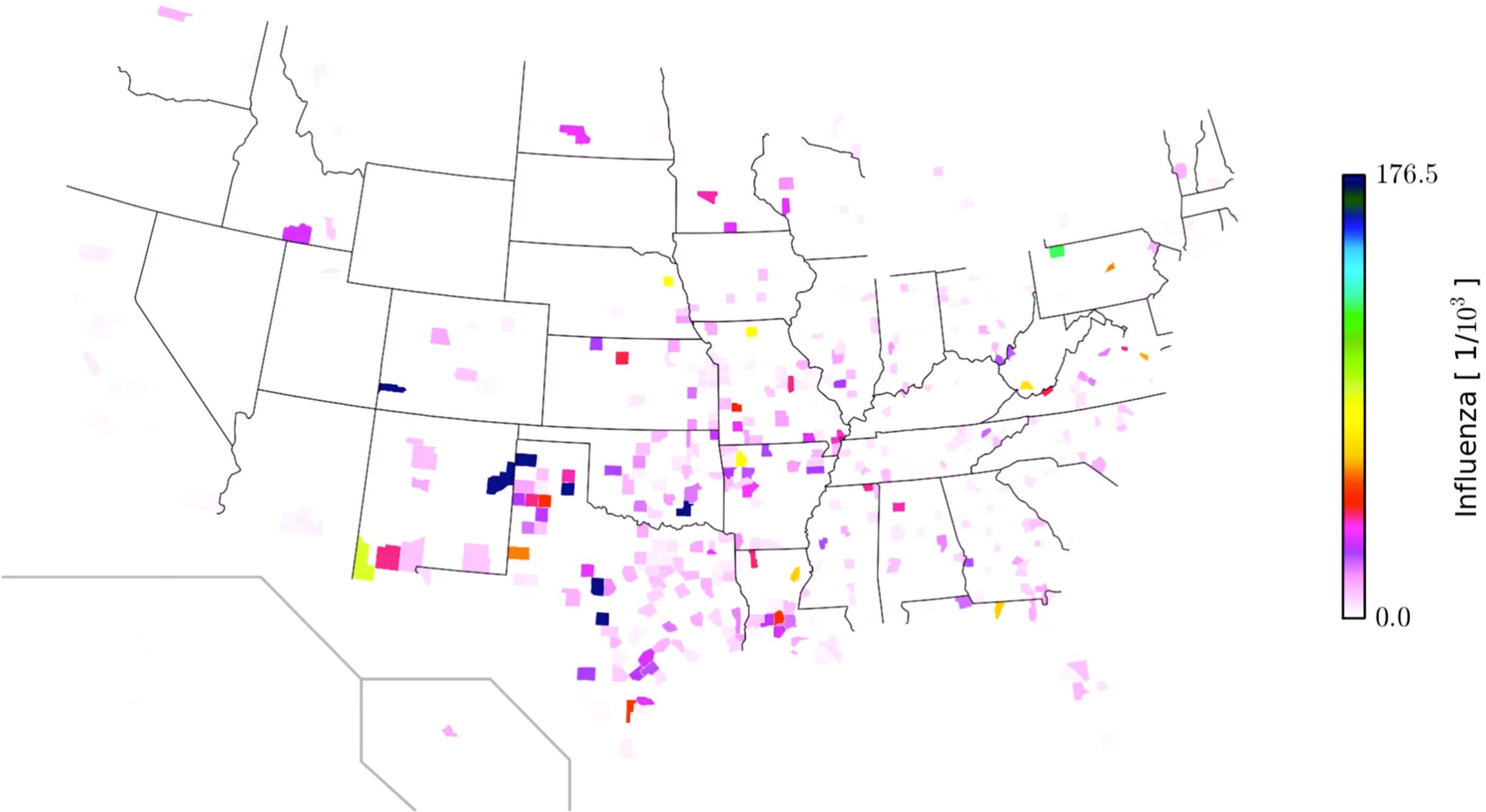


Data:

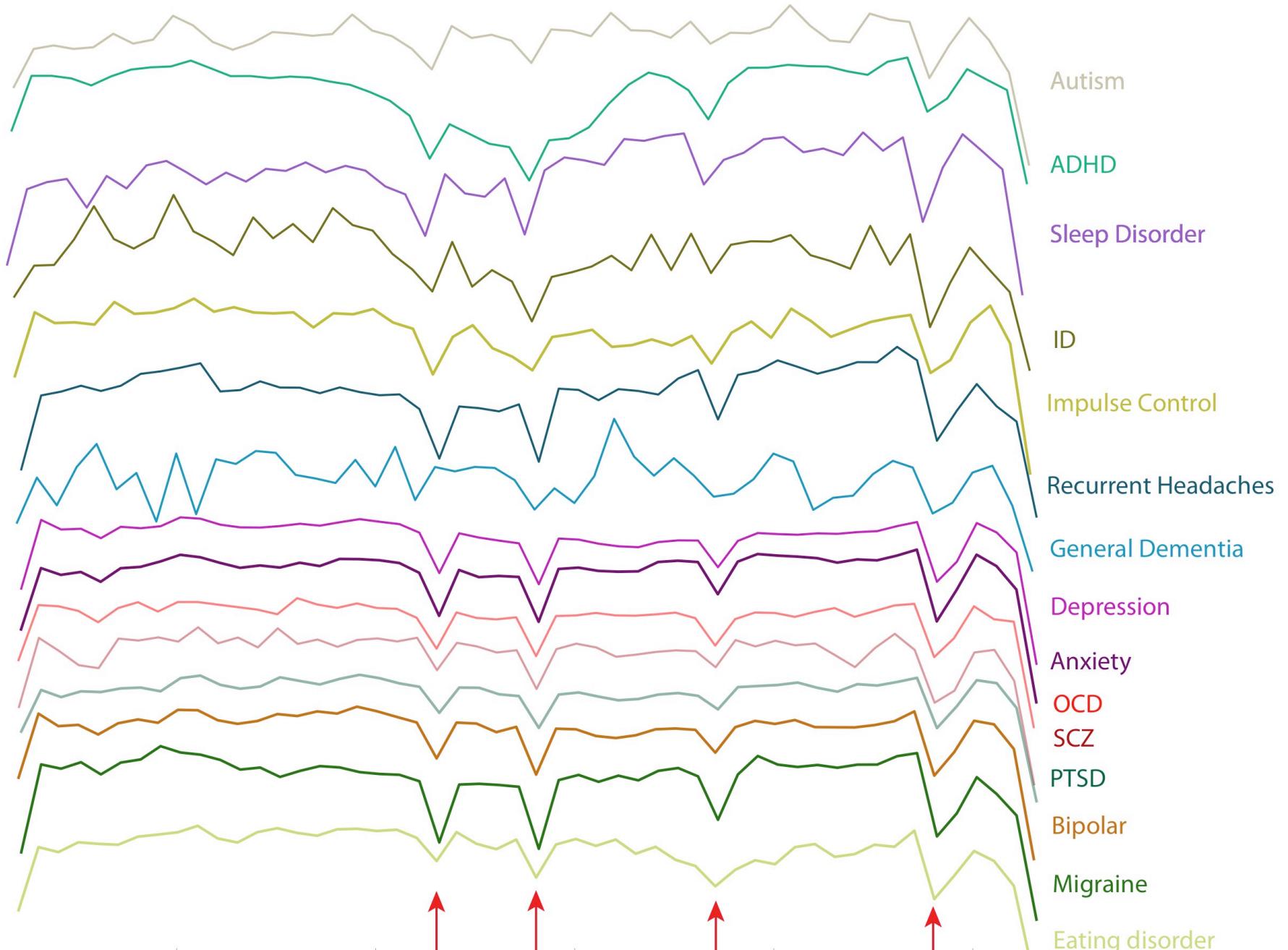
185,000,000 unique people  
USA, 16 years, day-level  
resolution

We have both time  
(diagnosis date) and  
space (US county  
equivalent) represented

Influenza 1

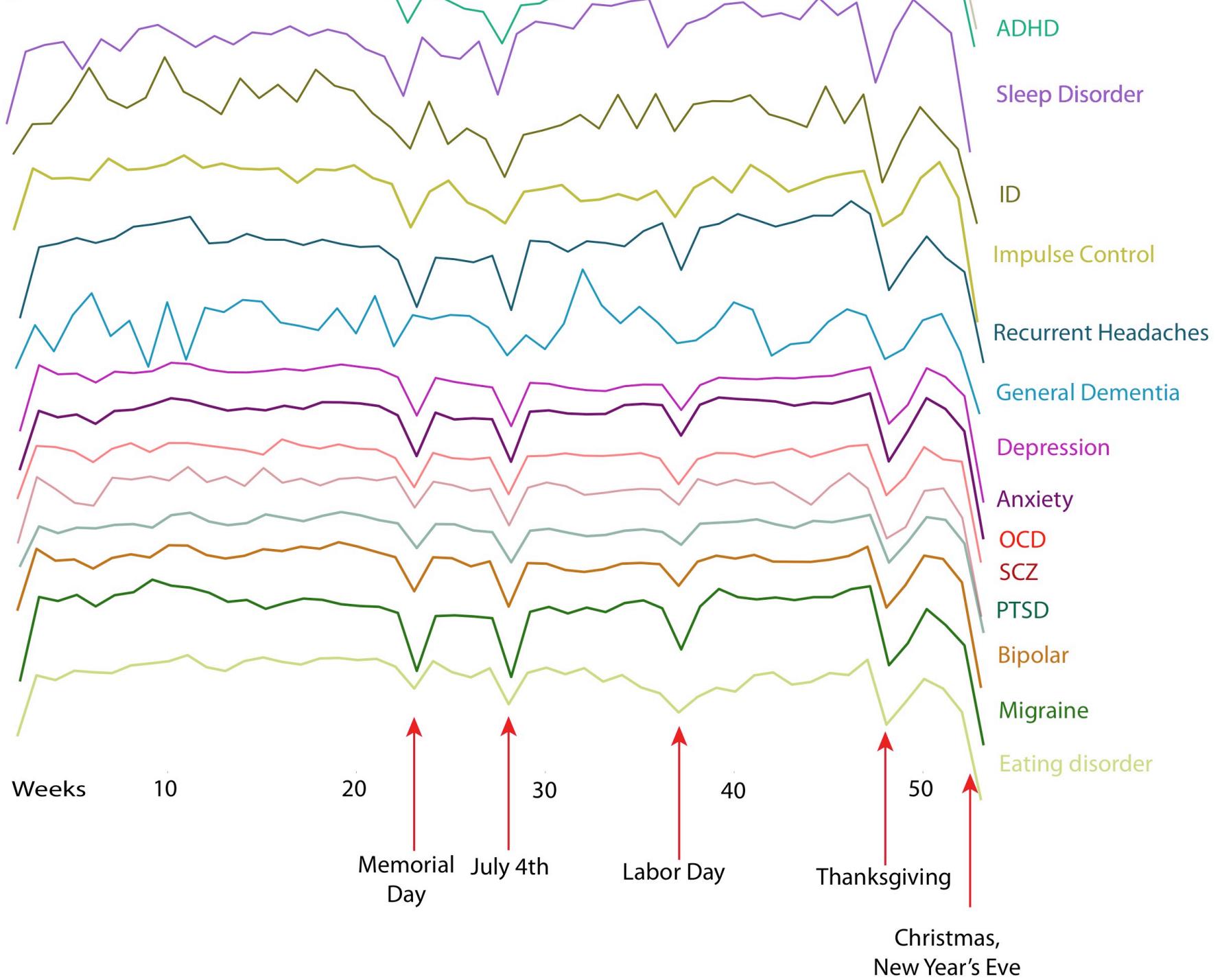


# Annual cycle of disease reporting



Weeks (1-52) →

Could you guess what these  
bumps and dips are?





A metaphor: etiology of a complex disease

# Charlotte, NC

# Air Midwest Flight 5481 crush

## January 8, 2003



# “Phenotype”

Stalled after takeoff, crashed into an US Airways hangar and burst into flames 37 seconds after leaving Charlotte/Douglas International Airport.

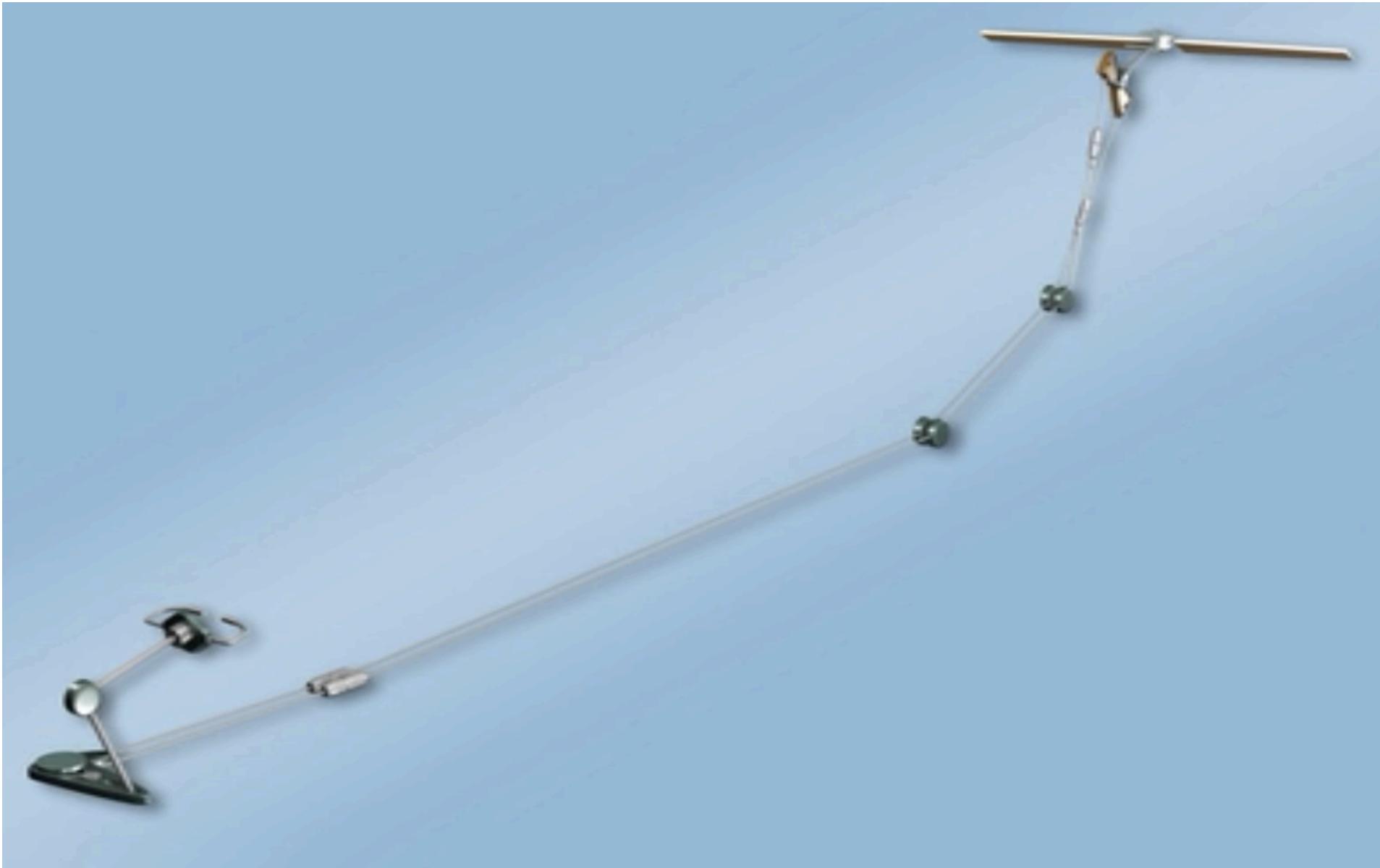
# Stalling plane



## Cause #1: “genetics”

...turnbuckles controlling tension on the cables to the elevators had been set incorrectly, ... pilots had insufficient pitch control.

# Elevator control cables



# Elevator control cables



## Cause #2: “environment”

... the plane was actually  
overloaded and out of balance.

# Outdated norms for the average weight:

The actual weight of an average passenger was more than 20 pounds (9 kg) greater than estimated.



The phenotype (the catastrophe) would not occur in the absence of either “genetic” or “environmental” component, both were required for the unfortunate outcome.

My point: it is insufficient to have only genetic or only environmental data for analysis of disease



# Environment

# Environment..





“ ... But there are also unknown unknowns -- the ones we don't know we don't know.”

Donald Rumsfeld

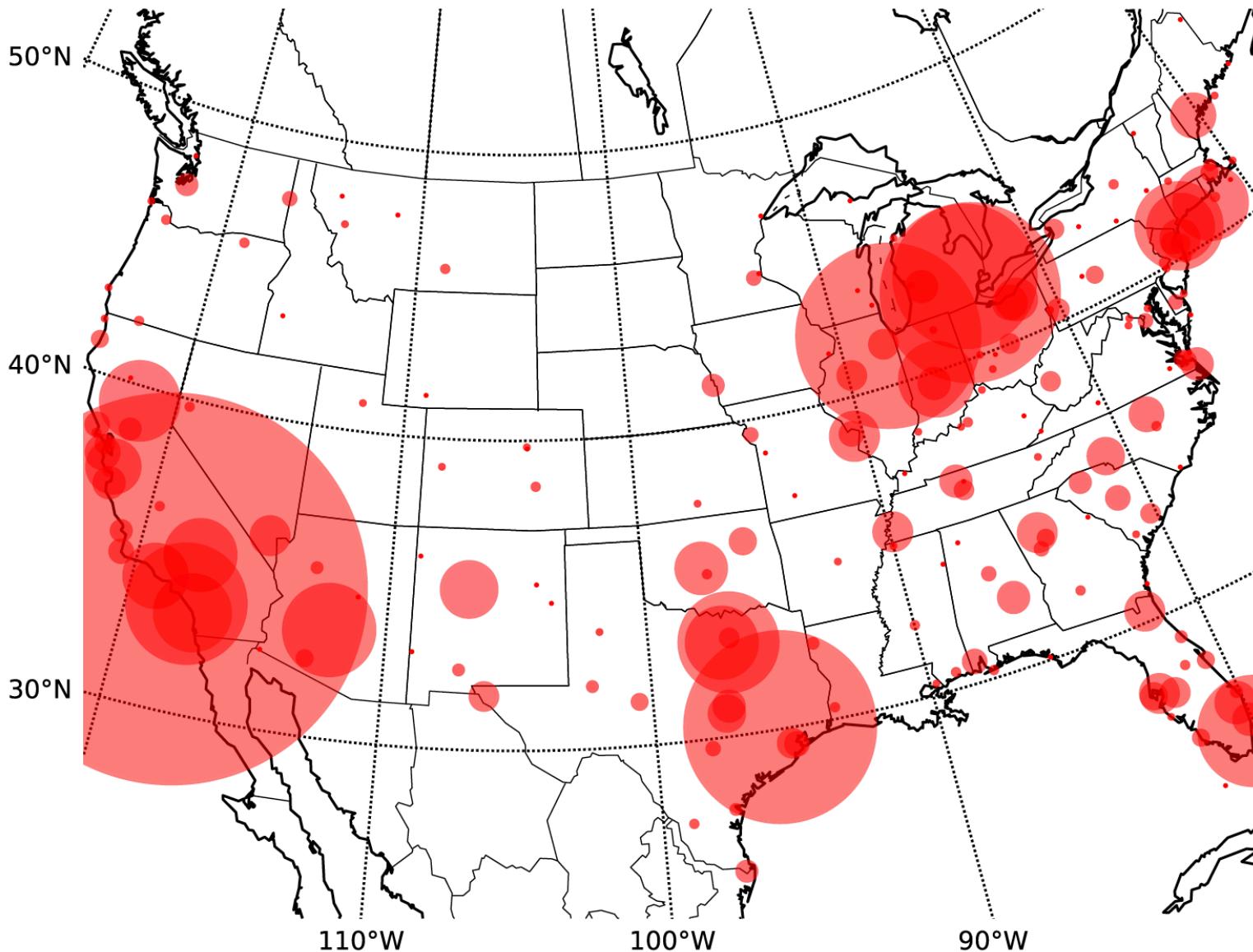
Millions (billions? Trillions?) of  
potentially relevant  
environmental factors

Trick for dealing with environment...

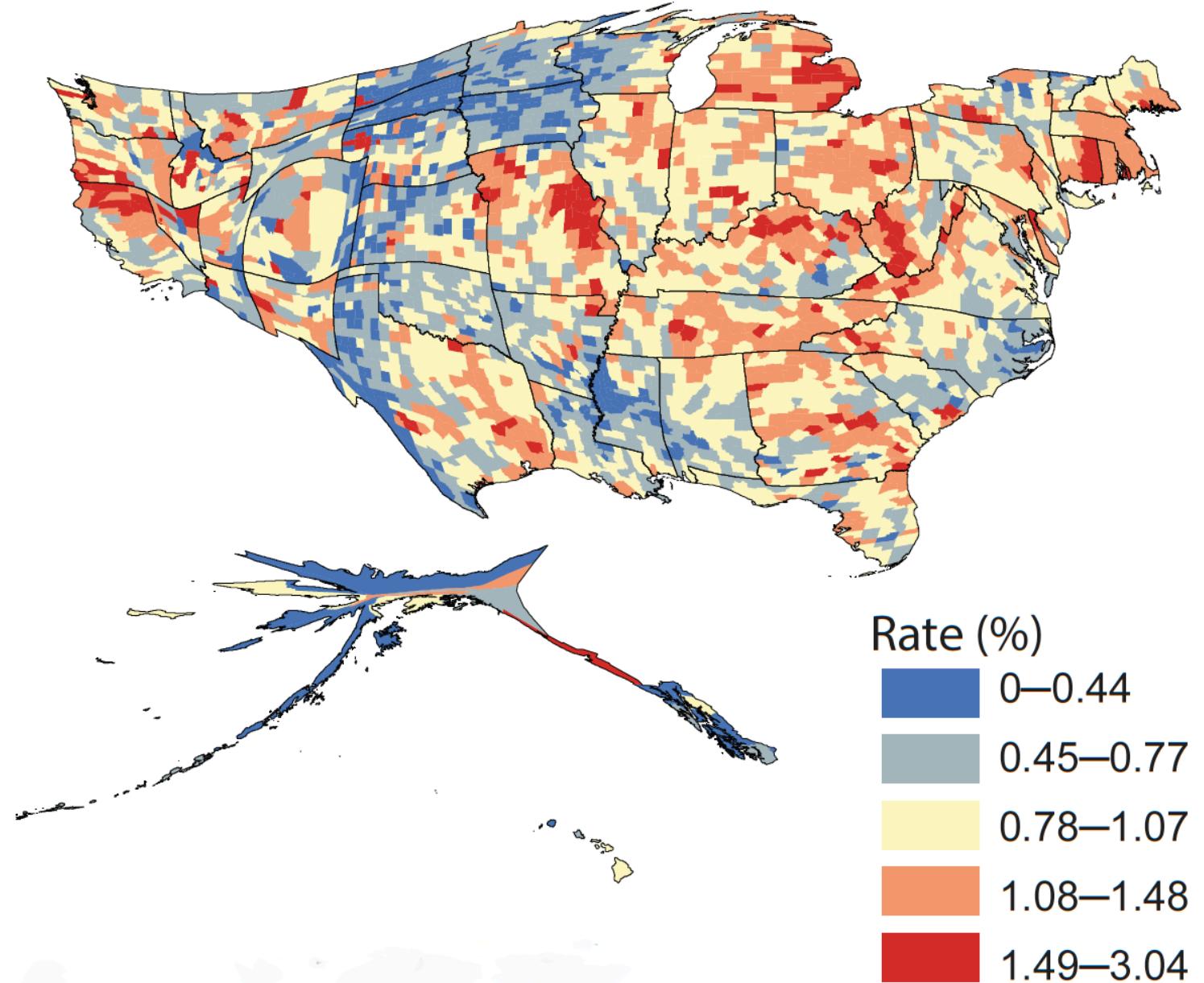
Diseases are not uniformly distributed over USA

# Idiopathic Pulmonary Fibrosis

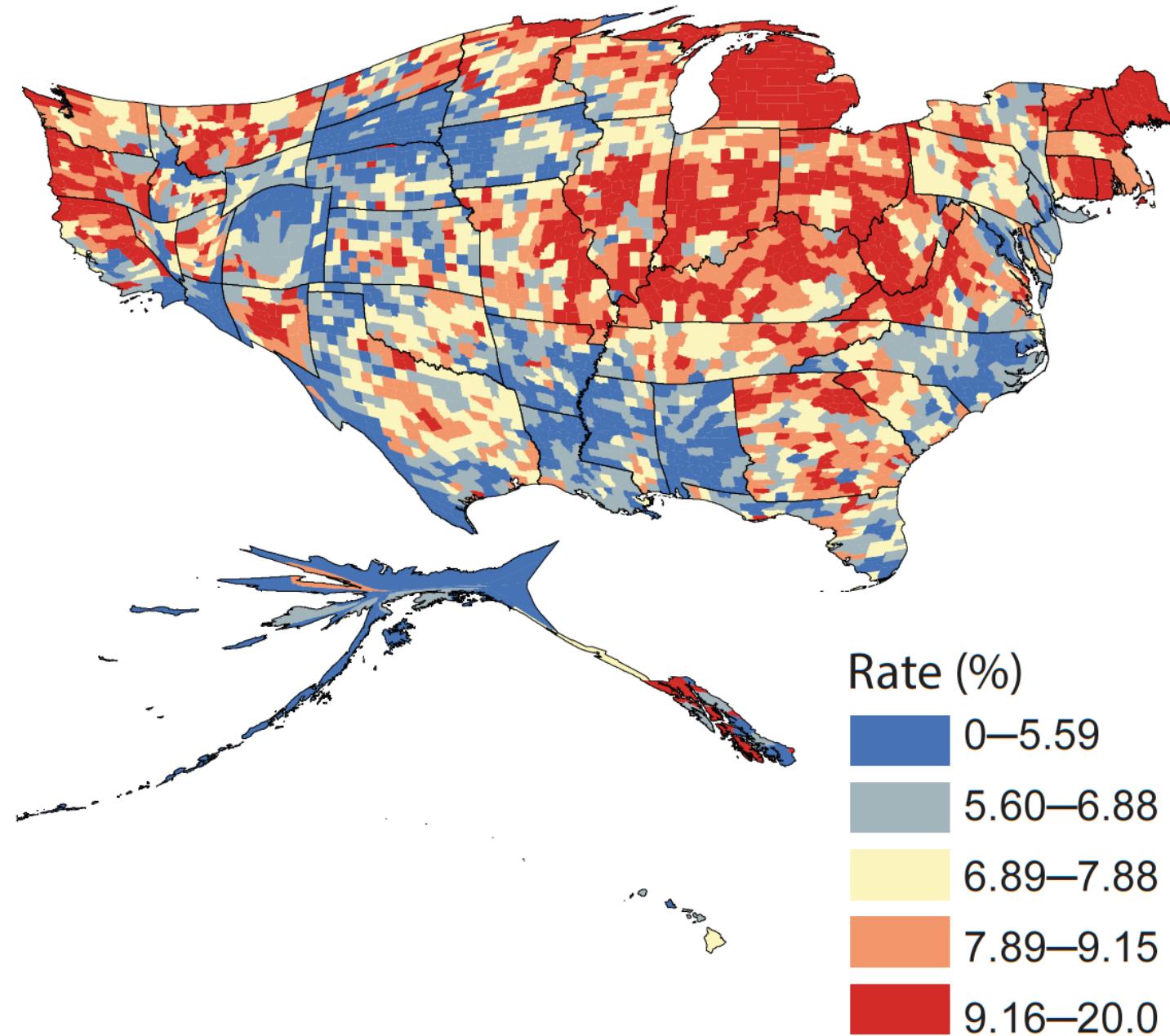
IPF



# Bipolar Disorder



# Major Depression





It is much harder to convince reviewers that  
**bipolar disease, depression, and**  
**schizophrenia** can be associated with  
**pollution**

**Environmental pollution is associated  
with increased risk of psychiatric disorders  
in US and Denmark**



**Environmental hack #1: pollution**

**Dr. Atif A. Khan**

[BROWSE](#)[PUBLISH](#)[ABOUT](#)[SEARCH](#)[advanced search](#)

OPEN ACCESS PEER-REVIEWED

SHORT REPORTS

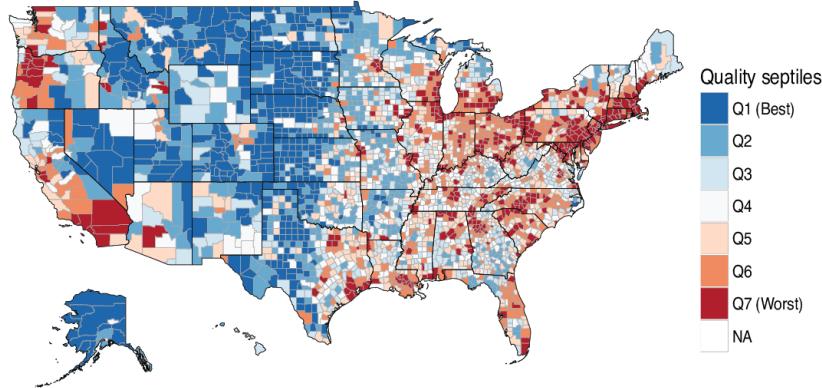
## Environmental pollution is associated with increased risk of psychiatric disorders in the US and Denmark

Atif Khan, Oleguer Plana-Ripoll, Sussie Antonsen, Jørgen Brandt, Camilla Geels, Hannah Landecker, Patrick F. Sullivan,  
Carsten Bøcker Pedersen, Andrey Rzhetsky

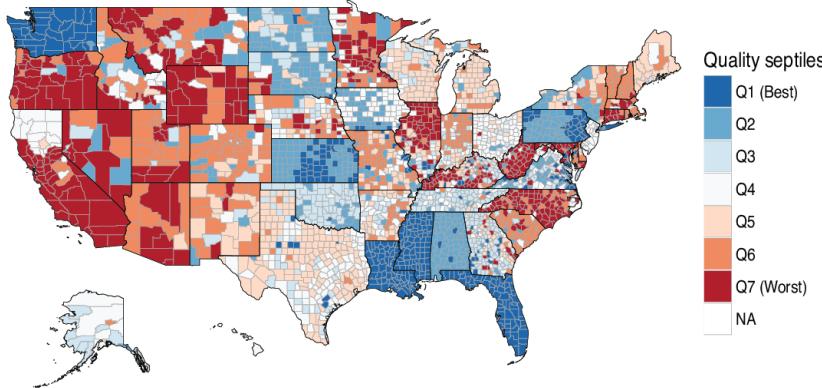
Published: August 20, 2019 • <https://doi.org/10.1371/journal.pbio.3000353>

<b>0</b> Save	<b>23</b> Citation
<b>39,695</b> View	<b>5</b> Share

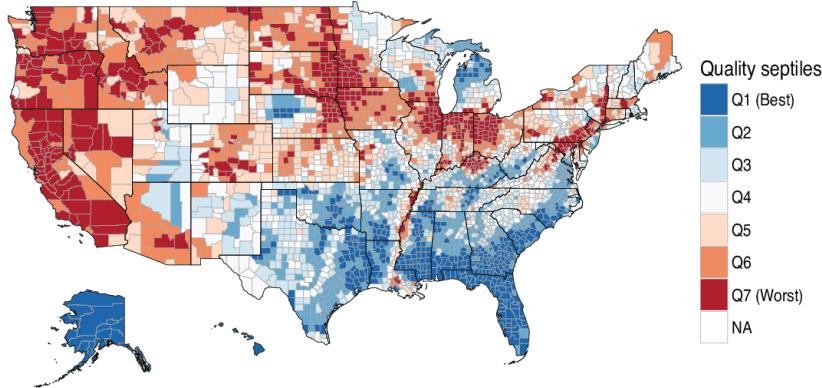
A: Air Quality



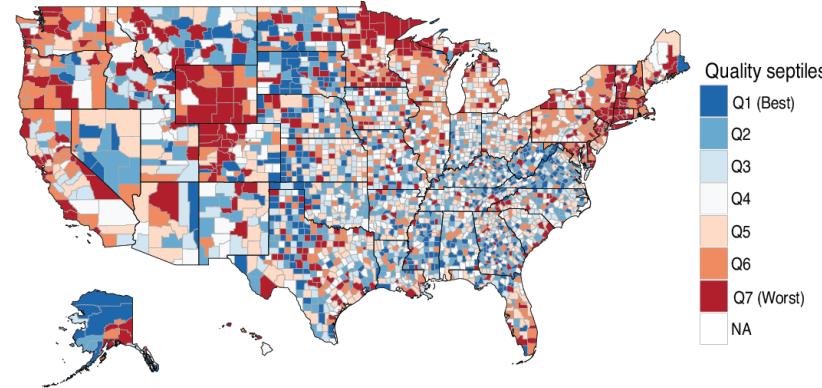
B: Water Quality



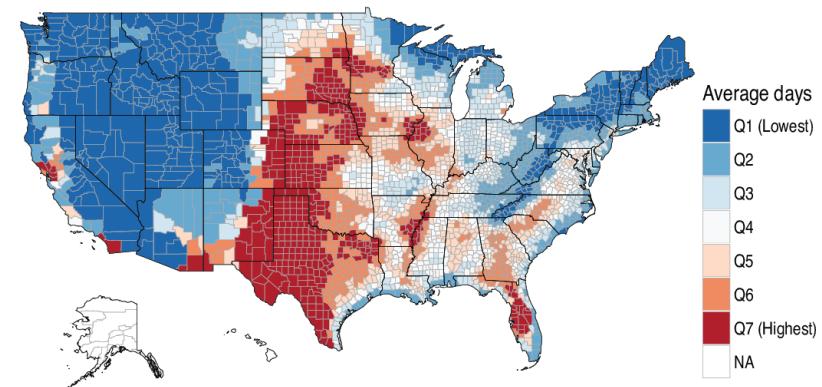
C: Land Quality



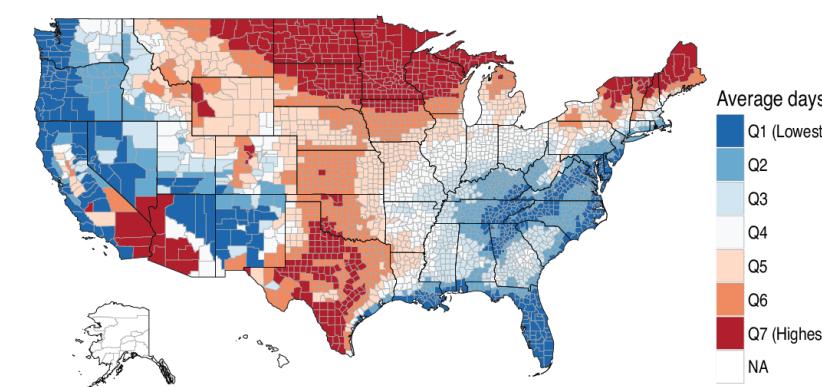
D: Built Quality



E: Good Weather Days



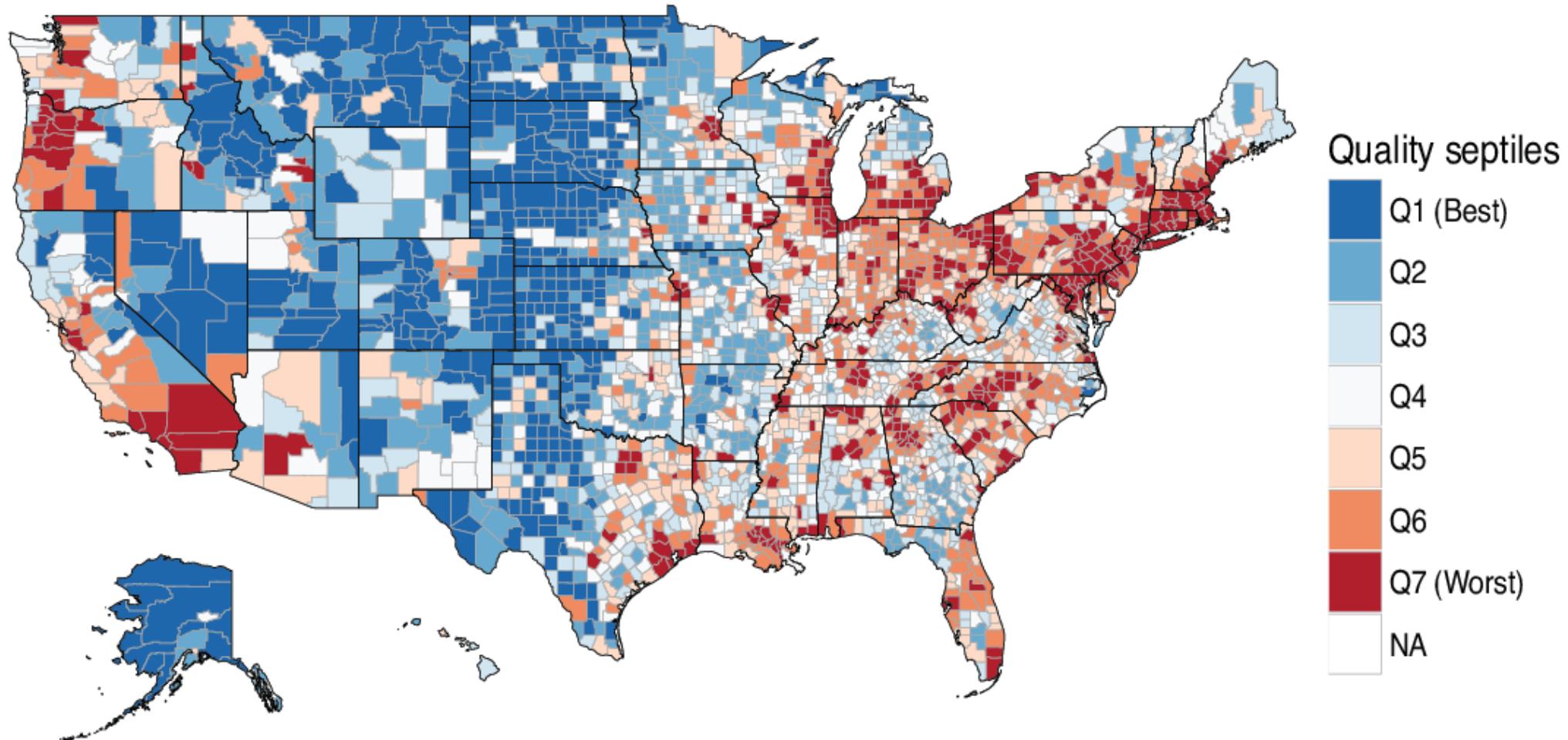
F: Bad Weather Days



US EPA  
environmental  
indices

# US EPA environmental indices

## A: Air Quality



# Mixed-effect regression

# Traces back to Ronald Aylmer Fisher



Fisher, RA (1918). "The correlation between relatives on the supposition of Mendelian inheritance". *Transactions of the Royal Society of Edinburgh* **52** (2): 399–433.

# Mixed-effect Poisson regression

$$\lambda = \beta_0 + \mathbf{B}^T \mathbf{x} + \mathbf{z}^T \mathbf{u}, \quad (1)$$

$$P(y|\lambda) = \frac{\lambda^y e^{-\lambda}}{y!}. \quad (2)$$

$y$  is non-negative count output variable,

$\lambda$  is non-negative Poisson rate parameter,

$\mathbf{x}$  is a predictor variable vector,

$\beta_0$  is an intercept parameter,

$\mathbf{b}$  is a vector of regression weight parameters

(fixed effect design),

$\mathbf{z}$  is random effects design vector,

$\mathbf{u}$  is random effects vector.

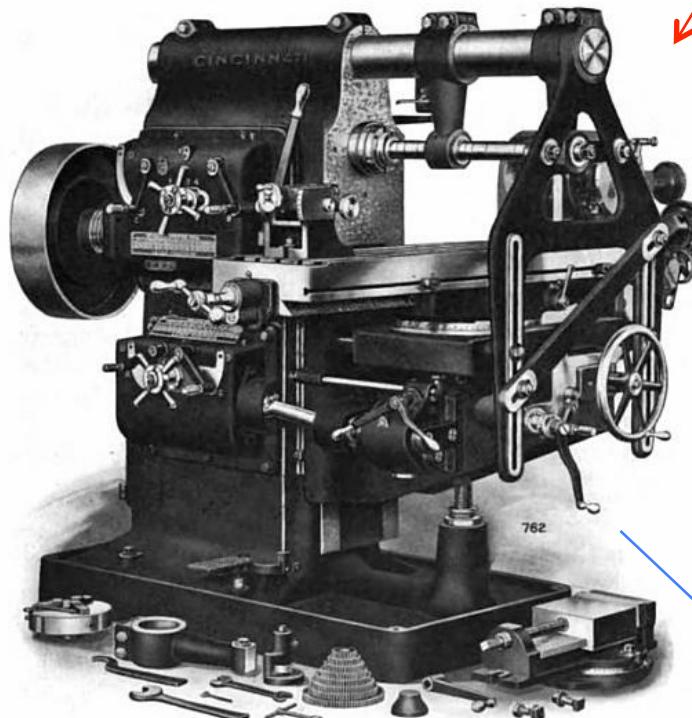
# Context:

*Fixed effects* = “known” genetics contribution from parental genotypes

*Random effects* = unknown (seemingly random) contributions from unknown genetics and environment

# Methods: Mixed-effect regression

$$\begin{aligned} \int f(x) dx &= \left( \sum_{j=1}^n a_j u_j(x) \right)' = \sum_{j=1}^n a_j u_j'(x) \\ f(x), \quad \left( \sum_{j=1}^n a_j u_j(x) \right)' &= \sum_{j=1}^n a_j u_j'(x) / x^{n+1} \xrightarrow{x \rightarrow 0} C = \lim_{x \rightarrow 0} f(x) \\ \Delta F = F(x_0 + \Delta x_0) - F(x_0) &= I_1 = \int \frac{1}{x} dx \xrightarrow{x \rightarrow 0} \infty \\ x_1 \pm y_1, \dots, x_n \pm y_n &= \left( \sqrt[n]{n+2} \right)^3 - \left( \sqrt[n]{n} \right)^3 \\ \sqrt[n]{n+2} &= \lim_{n \rightarrow \infty} \frac{\left( \sqrt[n]{n+2} \right)^3 - \left( \sqrt[n]{n} \right)^3}{\left( \sqrt[n]{n+2} \right)^2 + \left( \sqrt[n]{n+2} \right)} \sum_{k=0}^n a_k z^k \\ \left( 1 + \frac{1}{n+1} \right)^{n+1} &\leq \left( 1 + \frac{1}{n} \right)^n \xrightarrow{n \rightarrow \infty} a = \psi \left( \frac{1}{e} \right) \\ \int \pi f^2(x) dx &= \int \pi \left( \frac{r}{h} x \right)^2 dx = \int \frac{\pi r^2}{h^2} x^2 dx \int [u_1(x) + u_2(x)]^2 dx \\ \int x^2 \left[ \frac{7}{3} + \frac{3}{x} + \frac{5}{x^2} + \frac{1}{x^3} \right] dx &= P_n(z_0) = \sum_{k=0}^n a_k z^k = 0 \\ 4j \int f_j(x) dx + C &= \sum_{k=0}^n c_n^k n^{-k} x^k \int \left( \sum_{j=1}^n A_j \right) \\ z^{n-k} + a^k z^{n-k} + \dots + a^{n-k} &= I_1 = \int \frac{1}{x} dx \xrightarrow{x \rightarrow 0} \infty \\ a_0 + a_1 z + \dots + a_n z^n &= \sum_{k=0}^n a_k z^k \quad (a_k \neq 0) \quad P_n(z) = a \\ a(x+h) - \log_a x &= a = \psi \left( \frac{1}{e} \right) \quad (\log_a x) = \lim_{x \rightarrow 0} \frac{a(x+h) - a(x)}{h} \\ \lim_{h \rightarrow 0} \log_a \left( \frac{x+h}{x} \right)^{1/h} &= \lim_{h \rightarrow 0} \log_a \frac{1}{x} \left( 1 + \frac{h}{x} \right)^{1/h} \xrightarrow{h \rightarrow 0} \frac{1}{x} \log_a (1 + \frac{1}{x}) \\ \left( u_j(x) \right)' &= P_n(z_0) = \sum_{k=0}^n a_k z^k = 0 \quad I_1 = \int \frac{1}{x} dx \xrightarrow{x \rightarrow 0} \infty \quad \left( \sum_{j=1}^n a_j \right) \end{aligned}$$



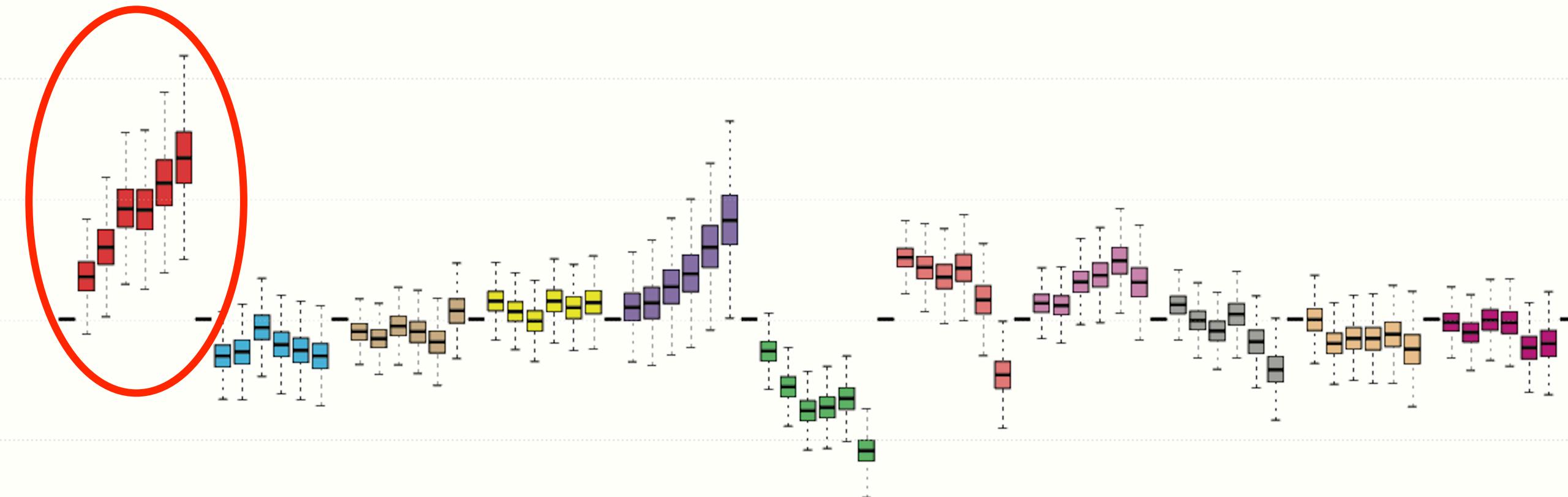
Factors

Effect on disease

# Results

Air Water Land Built Pop. Good Bad Insured Poor Urban Income  
Quality Quality Quality Quality Density Days Days

# Bipolar





Level of air pollution is associated with rates of bipolar disorder and depression in the USA.

With schizophrenia, bipolar disorder, personality disorder and depression in

Our results showed that in the United States, exposure to worst air quality was predictive of 27% and 6% increase in the rates of bipolar disorder and major depression respectively, when compared to the rates in best air quality regions.

Meanwhile, in Denmark, people exposed to highest air pollution had 162% increase in personality disorders, 148% in schizophrenia, 50% in major depression, and 29% in bipolar disorder when compared to those least exposed.

# Quite a bit of resonance...

Macht Luftverschmutzung psychisch krank?

23.8.2019 - 13:47, sda/dpa



Setzt verschmutzte Luft  
Bild: iStock

Nehmen nicht nur  
Schaden? Eine Studie  
die psychische G

SMOG UND PSYCHE  
Fördert schlechte Luft psychische Erkrankungen?

AKTUALISIERT AM 21.08.2019 - 15:00



Smog hat offenbar einen stärkeren Einfluss auf die Psyche als gedacht und

MEDIO AMBIENTE

## Vinculan la contaminación atmosférica al trastorno bipolar y la depresión

Un nuevo estudio se suma a las investigaciones que apuntan a la existencia de un vínculo entre la contaminación atmosférica y los problemas de salud mental.

miércoles, 21 de agosto de 2019



研究報告：汚濁空氣與患精神疾病有直接關係

2019年08月26日 11:30



Smog hat offenbar einen stärkeren Einfluss auf die Psyche als gedacht und

Ученые: возникновение психических расстройств связано с ухудшением экологии

Результаты соответствующего исследования были получены уже после поправок на возраст, пол, уровень дохода и другие обстоятельства жизни респондентов

СЕРГЕЙ КОЛЕСНИКОВ

Поделиться:  Комментарии:  11



The  
Guardian

led by readers



ion | Sport | Culture | Lifestyle | More ▾

Media Society Law Scotland Wales Northern Ireland

## Growing up in air-polluted areas linked to mental health issues

Correlation found between poor air quality and disorders including schizophrenia



Advertisement



The 5 Most Influential Visualizations of All T

GET THE WHITEPAPER





# Thank you!

