

# Data Visualization

I422

## Syllabus

[https://yongyeol.com/  
teaching/dviz\\_syllabus.pdf](https://yongyeol.com/teaching/dviz_syllabus.pdf)



# Intro



ronnie chen  
@rondoftw

day 3 with an e-bike and I have come to the conclusion that all the annoying evangelical e-bike enthusiasts weren't promoting them hard enough

11:11 AM · 26 Oct 22 · Twitter for Android

848 Retweets 134 Quote Tweets 10.6K Likes





ronnie chen  
@rondoftw

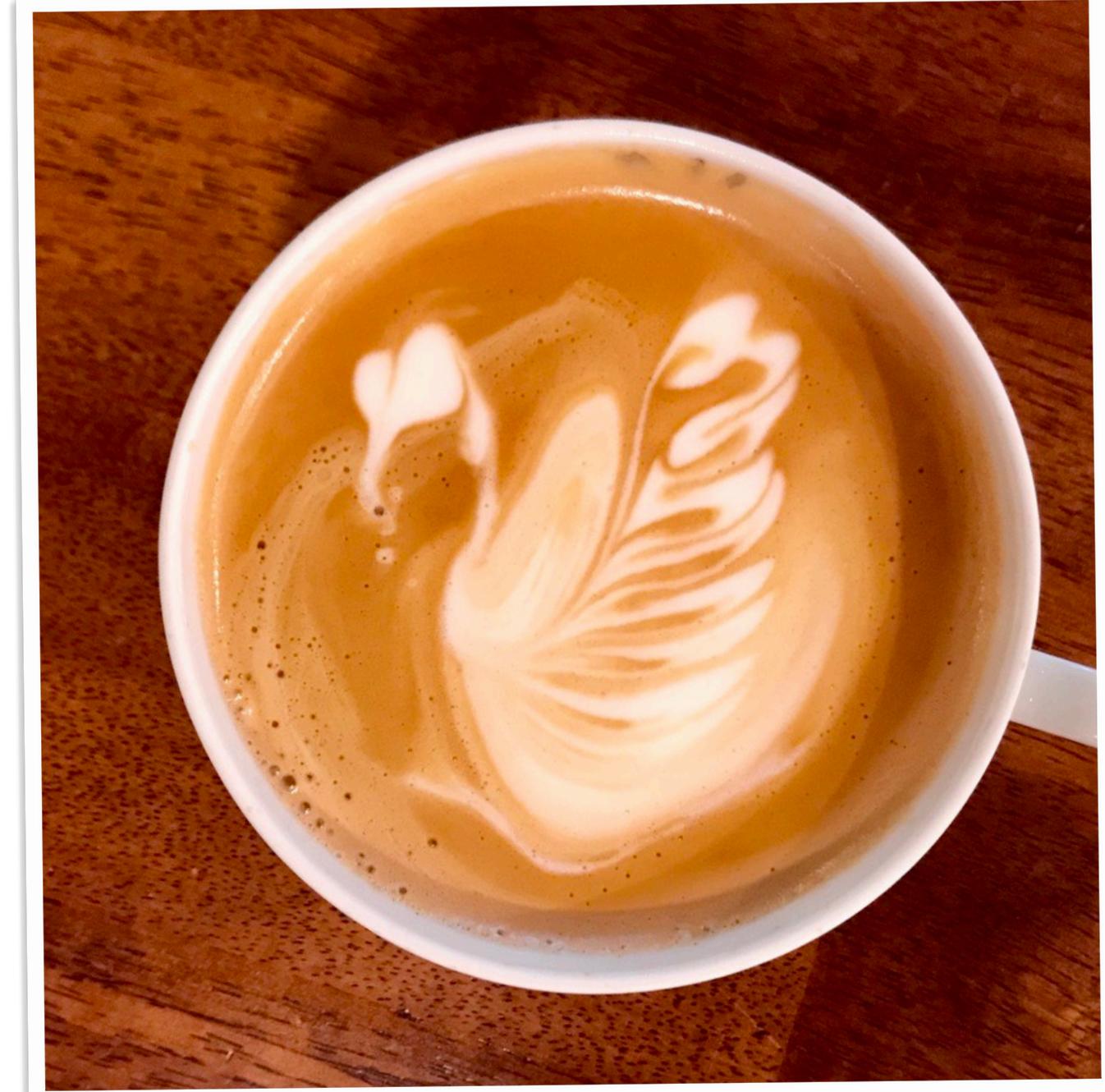
day 3 with an e-bike and I have come to the conclusion that all the annoying evangelical e-bike enthusiasts weren't promoting them hard enough

11:11 AM · 26 Oct 22 · Twitter for Android

848 Retweets 134 Quote Tweets 10.6K Likes



Yong-Yeol Ahn  
“YY”





ronnie chen  
@rondoftw

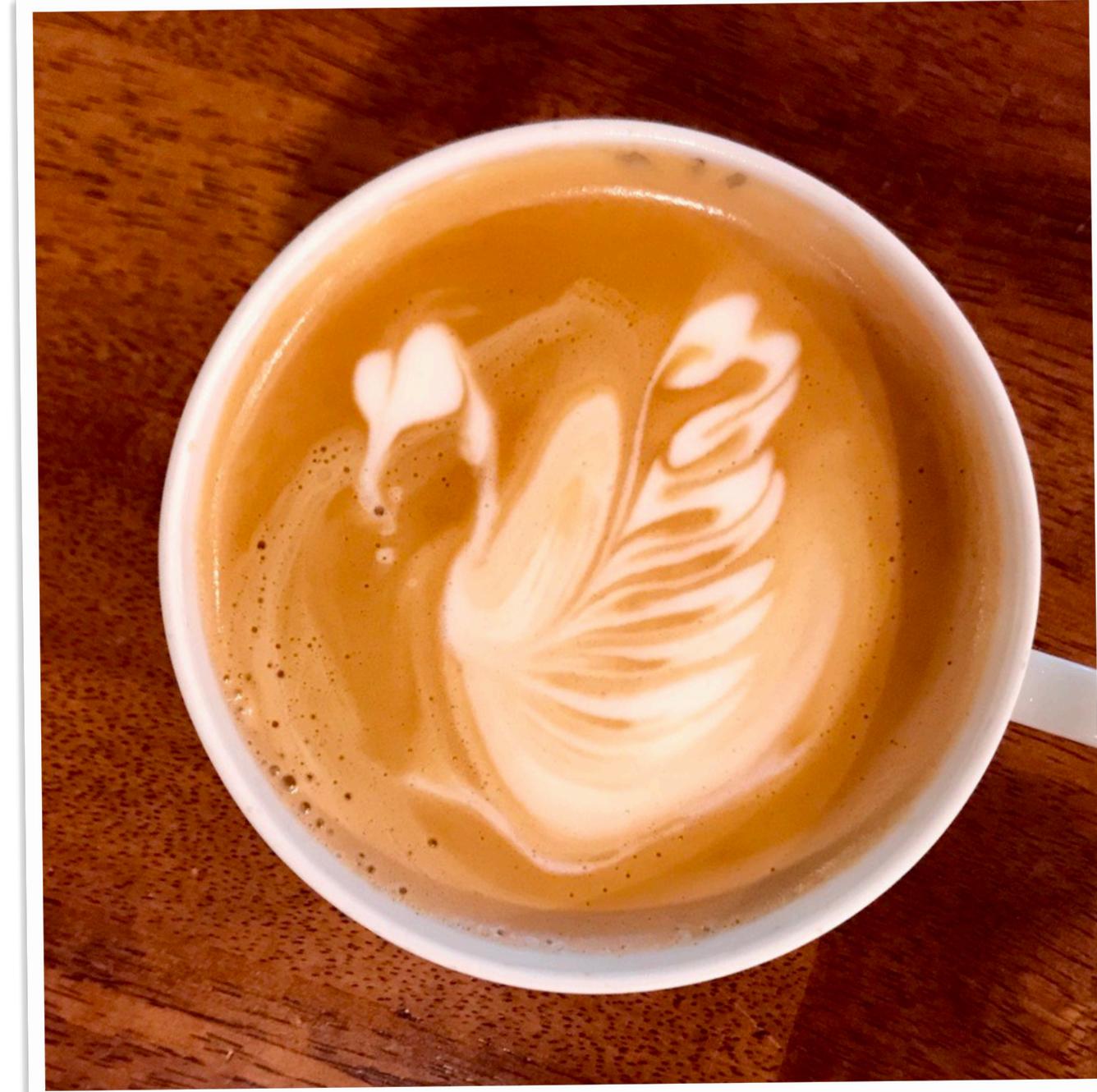
day 3 with an e-bike and I have come to the conclusion that all the annoying evangelical e-bike enthusiasts weren't promoting them hard enough

11:11 AM · 26 Oct 22 · Twitter for Android

848 Retweets 134 Quote Tweets 10.6K Likes



Yong-Yeol Ahn  
“YY”





ronnie chen  
@rondoftw

day 3 with an e-bike and I have come to the conclusion that all the annoying evangelical e-bike enthusiasts weren't promoting them hard enough

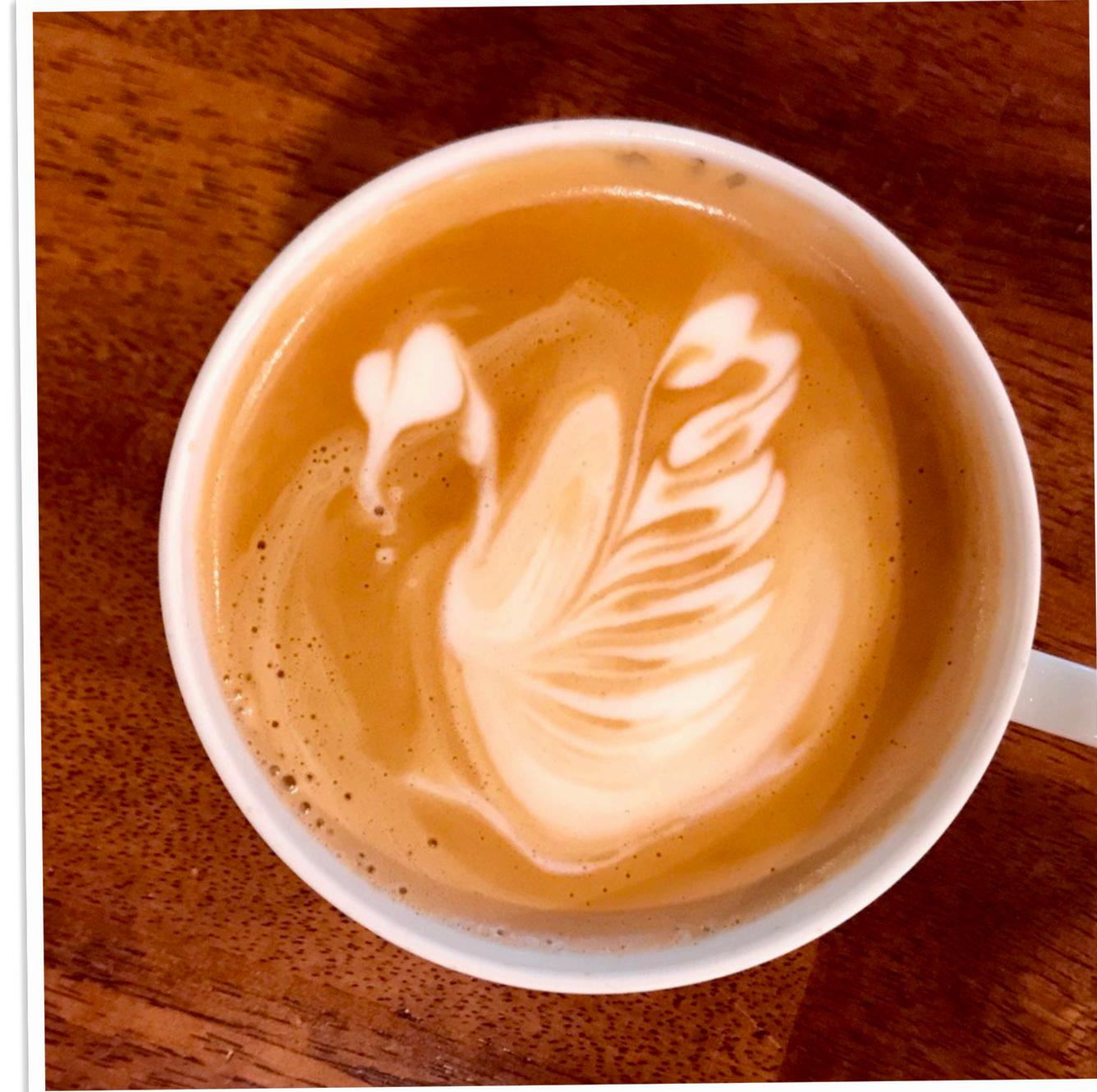
11:11 AM · 26 Oct 22 · Twitter for Android

848 Retweets 134 Quote Tweets 10.6K Likes



Yong-Yeol Ahn  
“YY”

yyahn@iu.edu





ronnie chen  
@rondoftw

day 3 with an e-bike and I have come to the conclusion that all the annoying evangelical e-bike enthusiasts weren't promoting them hard enough

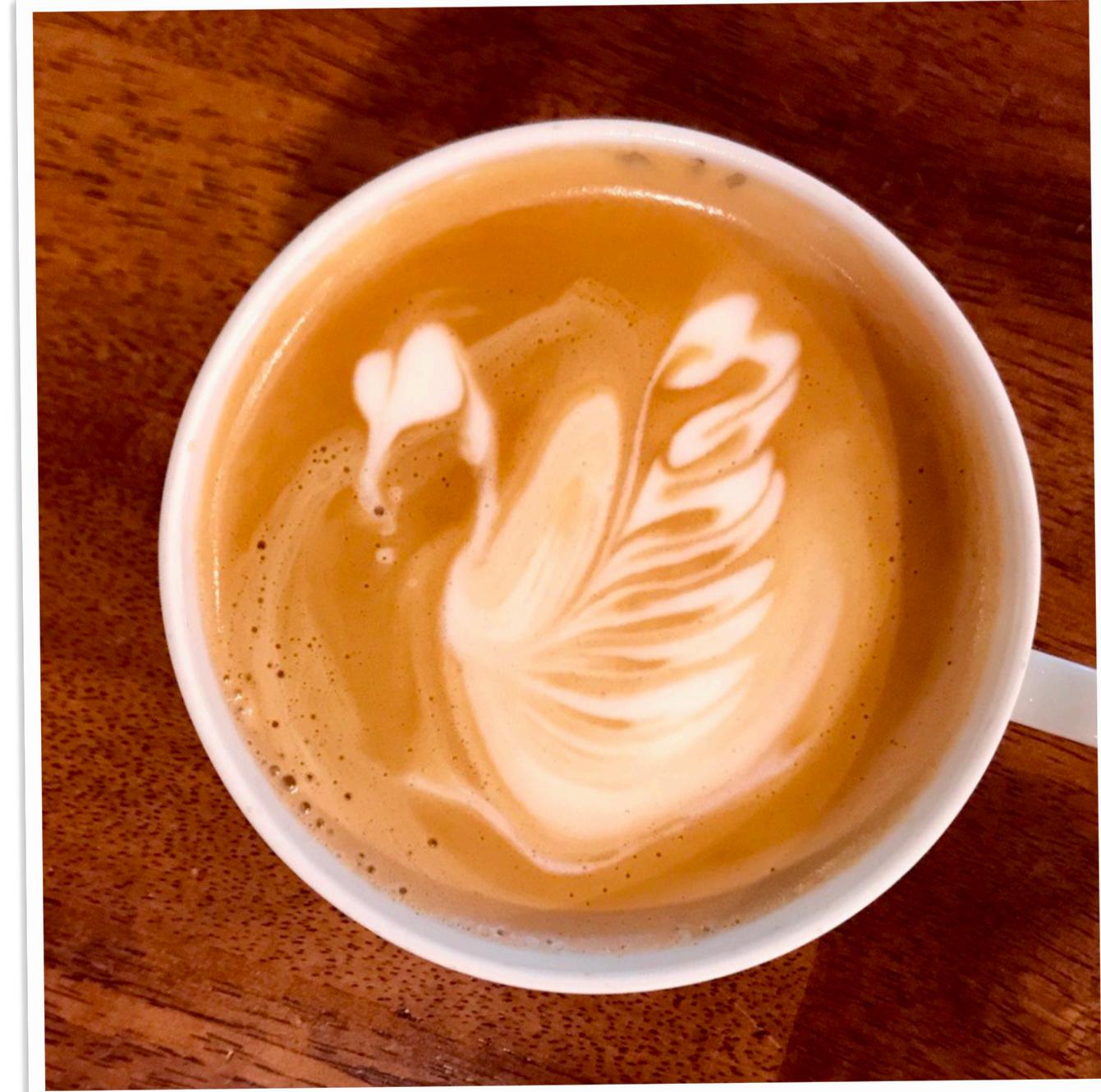
11:11 AM · 26 Oct 22 · Twitter for Android

848 Retweets 134 Quote Tweets 10.6K Likes



Yong-Yeol Ahn  
“YY”

[yyahn@iu.edu](mailto:yyahn@iu.edu)  
<http://yongyeol.com>





ronnie chen  
@rondoftw

day 3 with an e-bike and I have come to the conclusion that all the annoying evangelical e-bike enthusiasts weren't promoting them hard enough

11:11 AM · 26 Oct 22 · Twitter for Android

848 Retweets 134 Quote Tweets 10.6K Likes



Yong-Yeol Ahn  
“YY”

[yyahn@iu.edu](mailto:yyahn@iu.edu)  
<http://yongyeol.com>





ronnie chen  
@rondoftw

day 3 with an e-bike and I have come to the conclusion that all the annoying evangelical e-bike enthusiasts weren't promoting them hard enough

11:11 AM · 26 Oct 22 · Twitter for Android

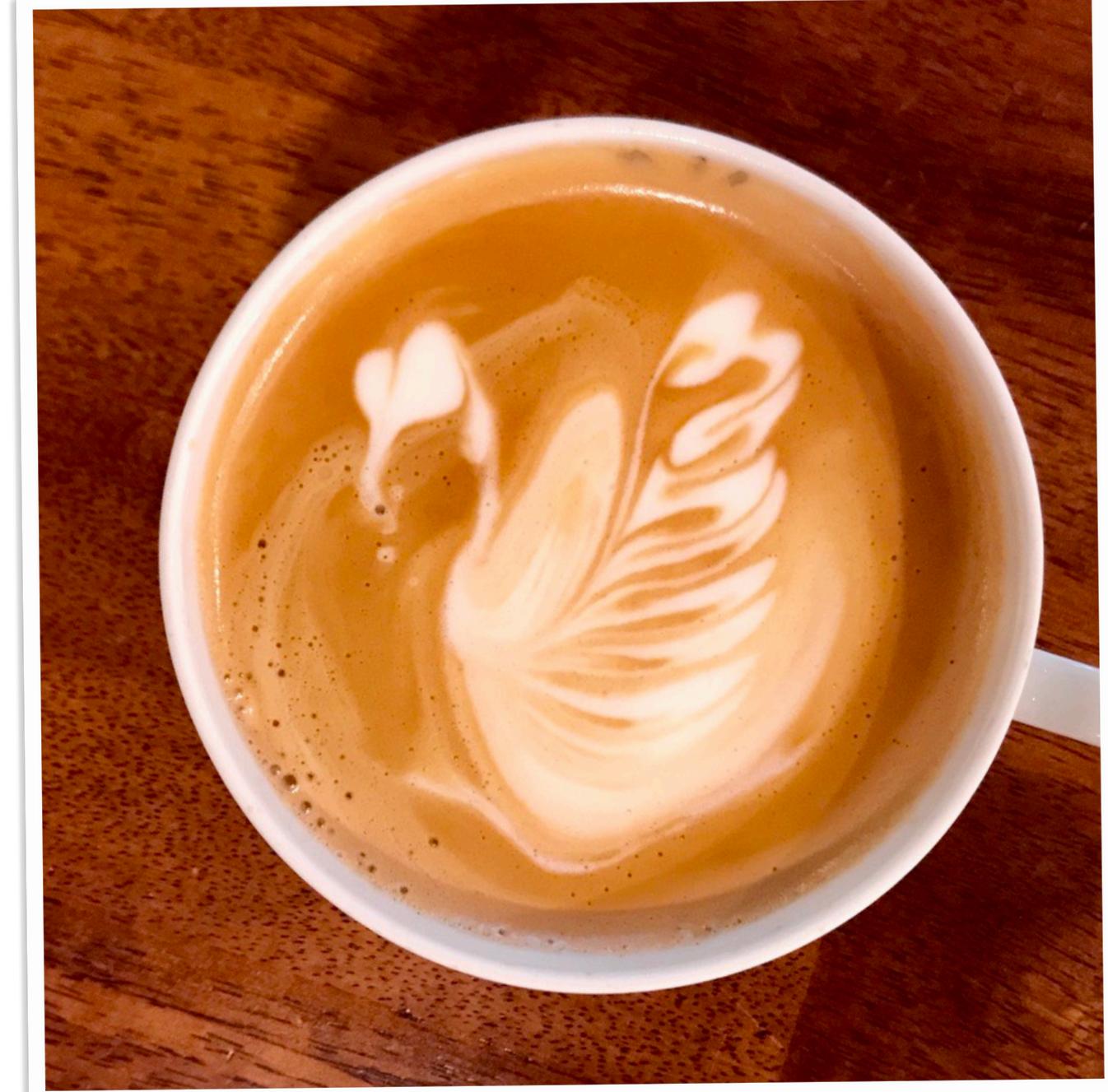
848 Retweets 134 Quote Tweets 10.6K Likes



Yong-Yeol Ahn  
“YY”

[yyahn@iu.edu](mailto:yyahn@iu.edu)  
<http://yongyeol.com>

안용열





ronnie chen  
@rondoftw

day 3 with an e-bike and I have come to the conclusion that all the annoying evangelical e-bike enthusiasts weren't promoting them hard enough

11:11 AM · 26 Oct 22 · Twitter for Android

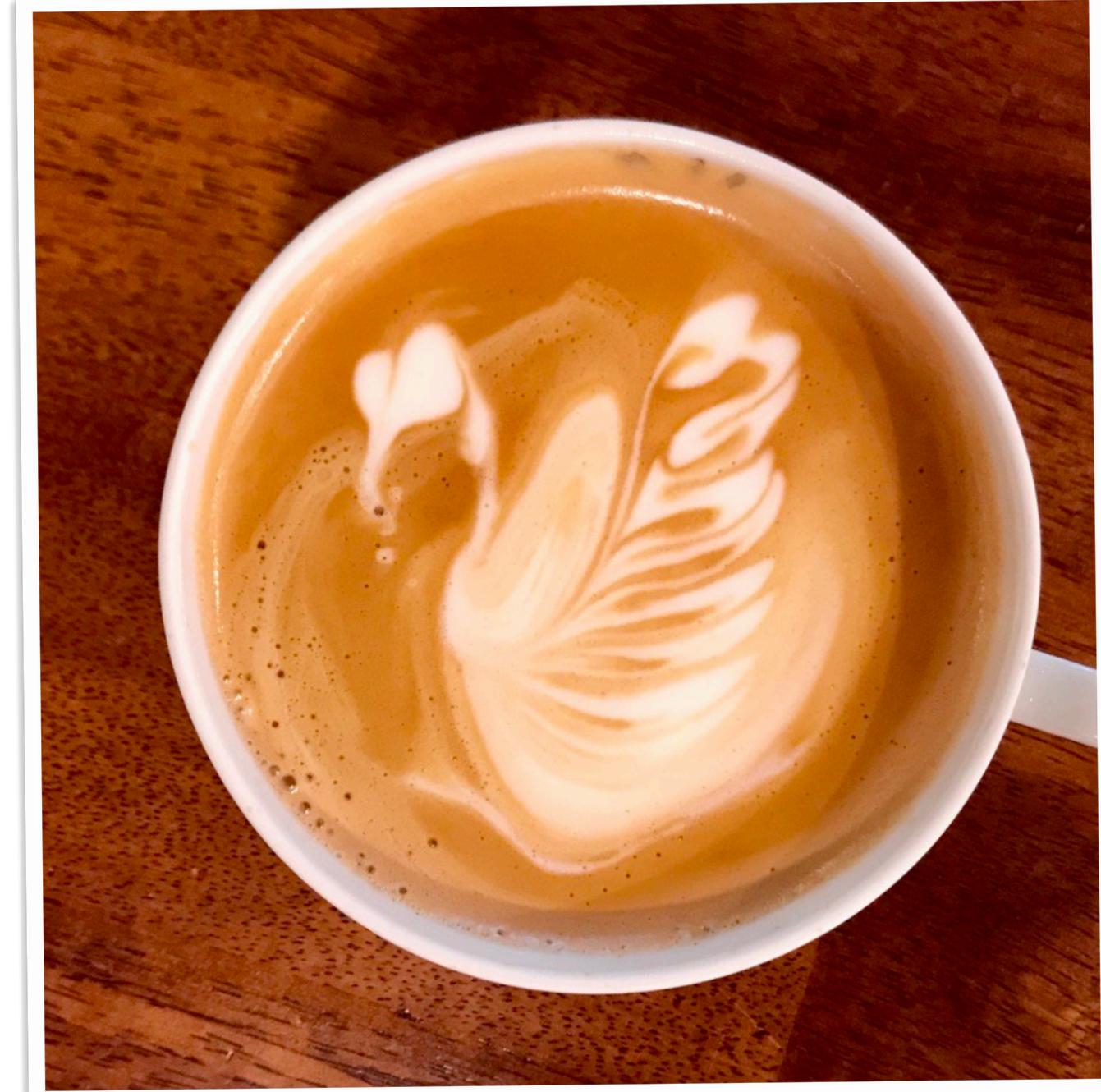
848 Retweets 134 Quote Tweets 10.6K Likes



Yong-Yeol Ahn  
“YY”

[yyahn@iu.edu](mailto:yyahn@iu.edu)  
<http://yongyeol.com>

안용열  
安用烈



# Assistant Instructors

Vincent Wong, [vmwong@iu.edu](mailto:vmwong@iu.edu)

Larry Zhang, [larzhang@iu.edu](mailto:larzhang@iu.edu)

Manasa Gudise, [mgudise@iu.edu](mailto:mgudise@iu.edu)

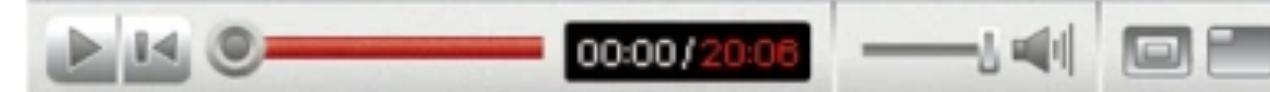
DECEMBER 25, 2006 / JANUARY 1, 2007

www.time.com

# TIME

PERSON OF THE YEAR

You.



Yes, you.

You control the Information Age.  
Welcome to your world.

# **Communication / Organization**

# Course Communication

Main: Canvas

Recommended for quicker

communication:  slack

<http://iu-dviz-course.slack.com/>

FALL 2018

Home

Modules

Announcements

Assignments

Discussions

Grades

People

IU Photo Roster

Files

Syllabus

Collaborations

SIS Performance  
Roster

SIS Grade Roster

Kaltura: My  
Media

Kaltura: Media  
Gallery

Quizzes

Outcomes

View Progress

+ Module

Module 0: Before we jump in

Complete All Items

Before we jump in

On motivation

View

Exciting misconceptions

View

Communication: Slack

View

Self-assessment questions

View

Honor pledge + Course survey participation

5 pts

Module 1: Why do we care?

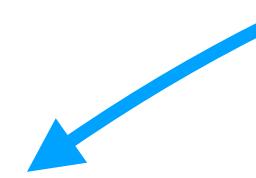
Complete All Items

Nice to meet you!

Introduce yourself!

# Course Activities

You can be excused if you notify us  
before the class begins!



- **Participation:** Attendance + quizzes (in-class and canvas), discussions, critiques
- **Assignments:** ~weekly assignments (mostly programming)
- **Project:** proposal, presentation, paper
- **Exam**

# Quiz & Discussion

- **In-class quizzes:** we will start each class with a quiz (from reading materials + past class materials). Why? *recalling* and *reconstructing* is better than listening or reading.
- **Canvas quizzes:** usually for checking your understanding of the canvas reading materials.
- **In class discussion:** Random discussion groups of 4-5 (each month)
- **Online discussion:** share your confusion or feedbacks through the **module recap** & ask questions on slack as well.
- **Visualization critique (canvas, once a semester):** pick an interesting visualization and write a critique on the techniques, design, messages, etc.

# Assignments

- Lab assignments (~each week, mostly programming-based)

# Projects

- Data analysis, Visualization tools, Explorable visualizations, ...
- <https://github.com/yy/dviz-course/wiki/Project-ideas>
- City of Bloomington data!

This course is end-heavy. So start thinking about the topic & team early

# Grades

- Quizzes, discussions, critiques: 20%
- Assignments: 20%
- Final exam: 30%
- Final projects: 30%
- (Final grades may be curved at the end)

**How to get the most out from  
the course**

# Motivation

Daniel Pink: Drive





# Multiple types of motives don't multiply the motivation of West Point cadets

Amy Wrzesniewski<sup>a,1</sup>, Barry Schwartz<sup>b</sup>, Xiangyu Cong<sup>c</sup>, Michael Kane<sup>c</sup>, Audrey Omar<sup>c</sup>, and Thomas Kolditz<sup>a</sup>

<sup>a</sup>Organizational Behavior, Yale School of Management, Yale University, New Haven, CT 06511; <sup>b</sup>Department of Psychology, Swarthmore College, Swarthmore, PA 19081; and <sup>c</sup>Yale Center for Analytical Sciences, Yale School of Public Health, Yale University, New Haven, CT 06520

Edited\* by Linda M. Bartoshuk, University of Florida, Gainesville, FL, and approved June 4, 2014 (received for review March 29, 2014)

Although people often assume that multiple motives for doing something will be more powerful and effective than a single motive, research suggests that different types of motives for the same action sometimes compete. More specifically, research suggests that instrumental motives, which are extrinsic to the activities at hand, can weaken internal motives, which are intrinsic to the activities at hand. We tested whether holding both instrumental and internal motives yields negative outcomes in a field context in which various motives occur naturally and long-term educational and career outcomes are at stake. We assessed the impact of the motives of over 10,000 West Point cadets over the period of a decade on whether they would become commissioned officers, extend their officer service beyond the minimum required period, and be selected for early career promotions. For each outcome, motivation internal to military service itself predicted positive outcomes; a relationship that was negatively affected when instrumental motives were also in evidence. These results suggest that holding multiple motives damages persistence and performance in educational and occupational contexts over long periods of time.

crowding out | overjustification effect

Philosophers and psychologists have long distinguished among various types of motives to engage in particular activities. Some motives are internal or intrinsic to the activities themselves. The artist is motivated to create a great painting. The

activities need not be undertaken for the inherent pleasure they bring (2, 3).

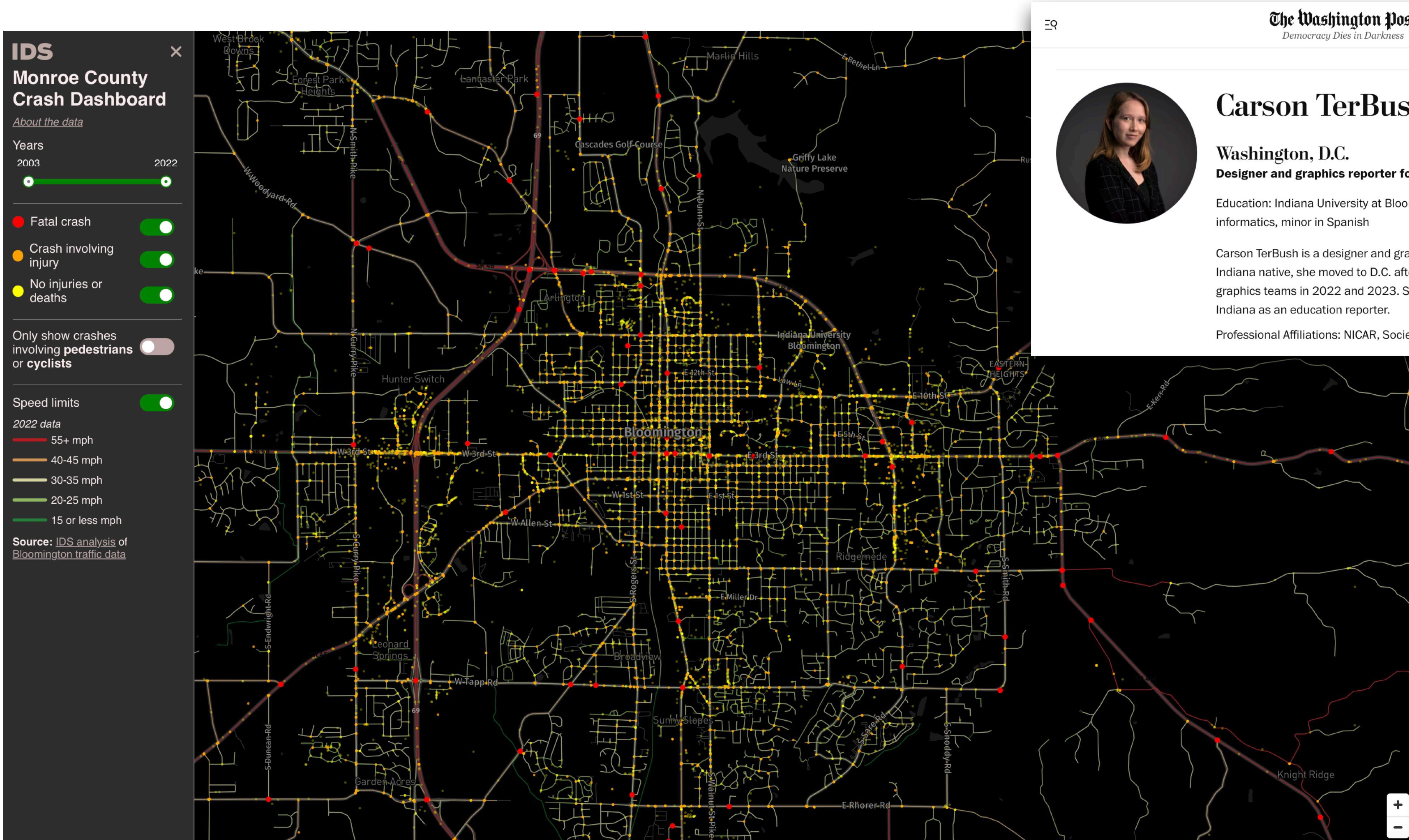
What happens to the performance of demanding, effortful activities when internal and instrumental motives are combined? Logic would suggest that if you have one reason for doing something, having two or more reasons to do the same thing would be even better, rendering motivation more tenacious, follow-through stronger, and outcomes better. Schools and workplaces are full of systems that attempt to tap people's internal motives to act (e.g., because engaging in the activity is the moral, interesting, or meaningful thing to do), while also providing rewards intended to spark instrumental motives to pursue the same acts (e.g., grades, bonuses, promotions, and so forth). Tapping internal motives and the instrumental motives assumed to result from rewards seems to be a foolproof way to engage the full spectrum of motivational levers that lead people to act. However, in a direct challenge to this assumption, social science research suggests that far from boosting motivation, holding instrumental motives can undermine whatever internal motives may have been operating, leading to drops in overall motivation, persistence, and performance (4–6). In short, this work suggests that salient instrumental incentives trigger instrumental motives, acting to undermine motivation that would otherwise be based in the value and reward of doing the activity or engaging in the act for the sake of objectives that are intimately connected to the act itself. This effect, labeled the "motivational crowding out effect" by economists (7) and the "overjustification effect" by psycho-

**FORGET** about the grades

# **Opportunities**

## **Mastery**

## **Fun**



**What this course can do:  
showing the starting points**

**“College is a place where a professor’s lecture notes go straight to  
the students’ lecture notes,**

***without passing through the brains of either.***

–Anon. (Often wrongly attributed to Mark Twain)

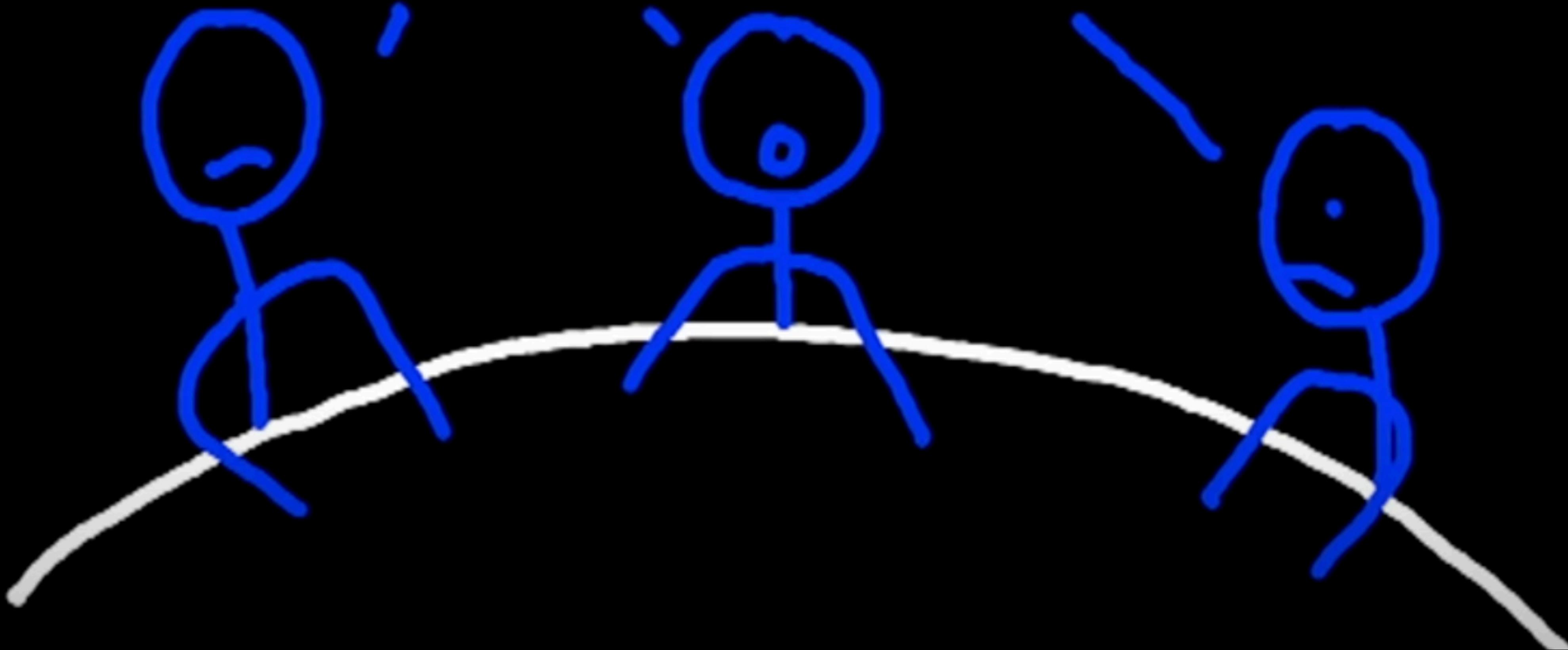


<https://www.youtube.com/watch?v=RQaW2bFieo8>



<https://www.youtube.com/watch?v=RQaW2bFieo8>

CONFUSING!





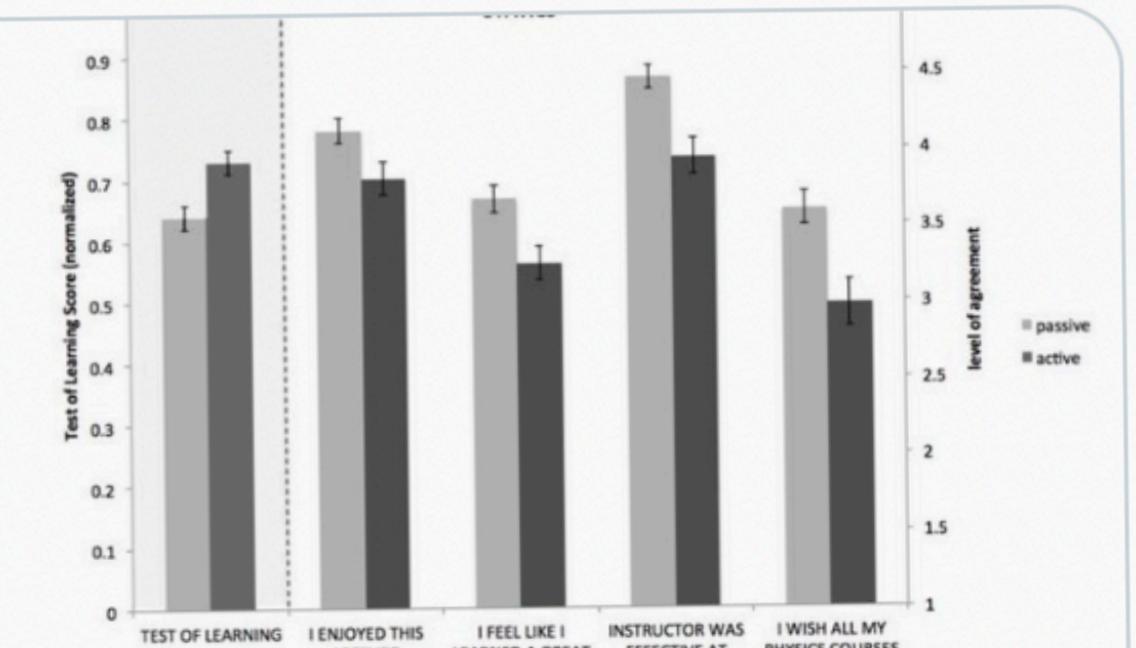
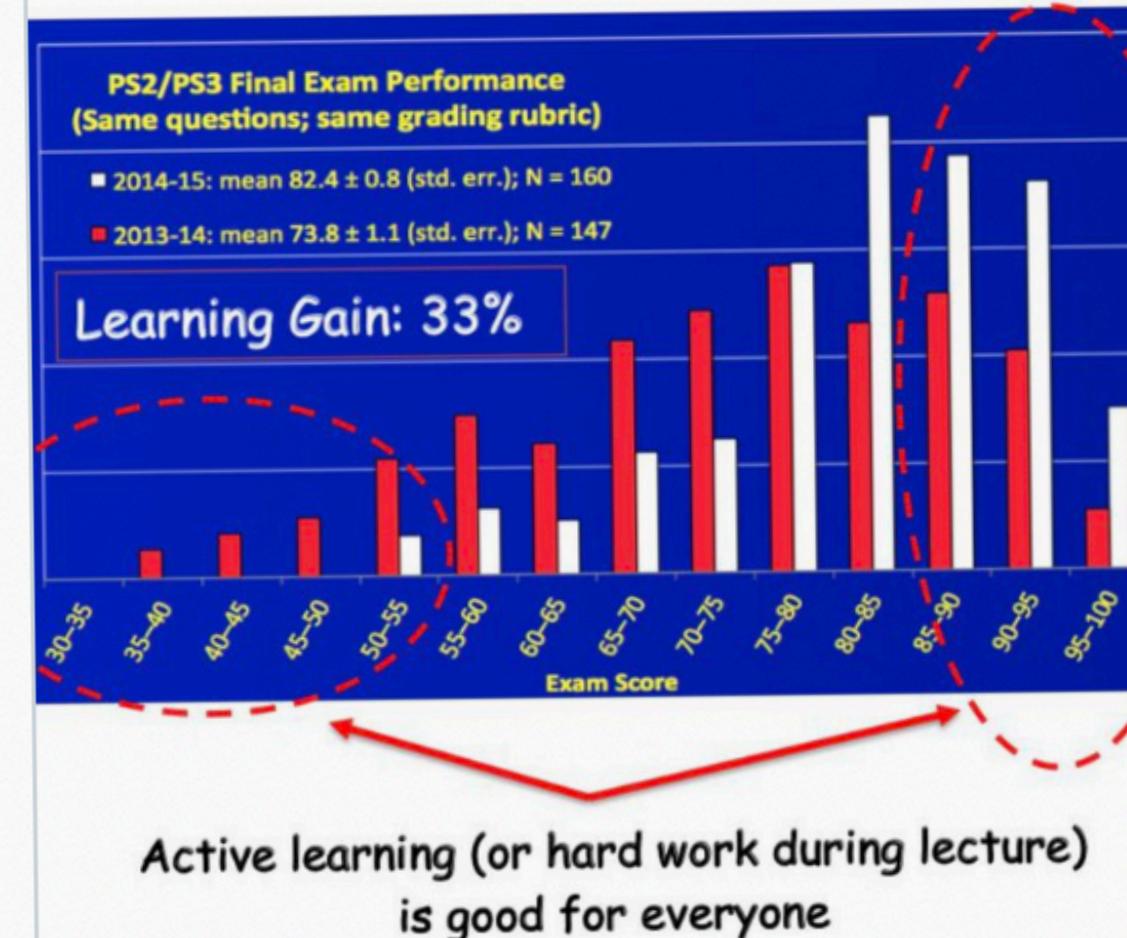
Ethan Mollick   
@emollick

...

人群中的一员 An experiment at Harvard showed active learning upped test scores by 33%... but students thought they were learning more from non-active lectures. The problem: being challenged is the key to learning BUT we hate challenges, they are hard & make you realize how little you know.

slides for 20-minute-long presentation given by the instructor at the beginning of the semester aimed at improving student attitudes about active learning:

### Results: Student Performance on Identical 3-hour Final Exam Questions in this Course



no effort to persuade students of the benefit of either method. Students in active classrooms learned more (as would be expected based on prior research), but their perception of learning, while positive, was lower than that of their peers in passive environments. This suggests that attempts to evaluate and reward student learning in active classrooms could inadvertently promote inferior (passive) pedagogical methods. For instance, a superstar lecturer could create such a positive feeling of learning that students would choose those lectures over active learning. Most importantly, these results suggest that when students experience the increased cognitive effort associated with active learning, they initially take that effort to signify poorer learning. That disconnect may have a detrimental effect on students' motivation, engagement, and ability to self-regulate their own learning. Although students can, on their own, discover the increased value of learning actively engaged during a semester, cognitive learning may be impaired during the initial part of the course. We discuss strategies that instructors can use, early in the semester, to improve students' response to being actively engaged in the classroom.

**Materials and Methods**  
Our study sought to measure students' perception of learning when active learning alone is toggled on and off. This contrasts with typical educational interventions that include active engagement as one component of many changes to a course. We compared actual learning to students' feeling of

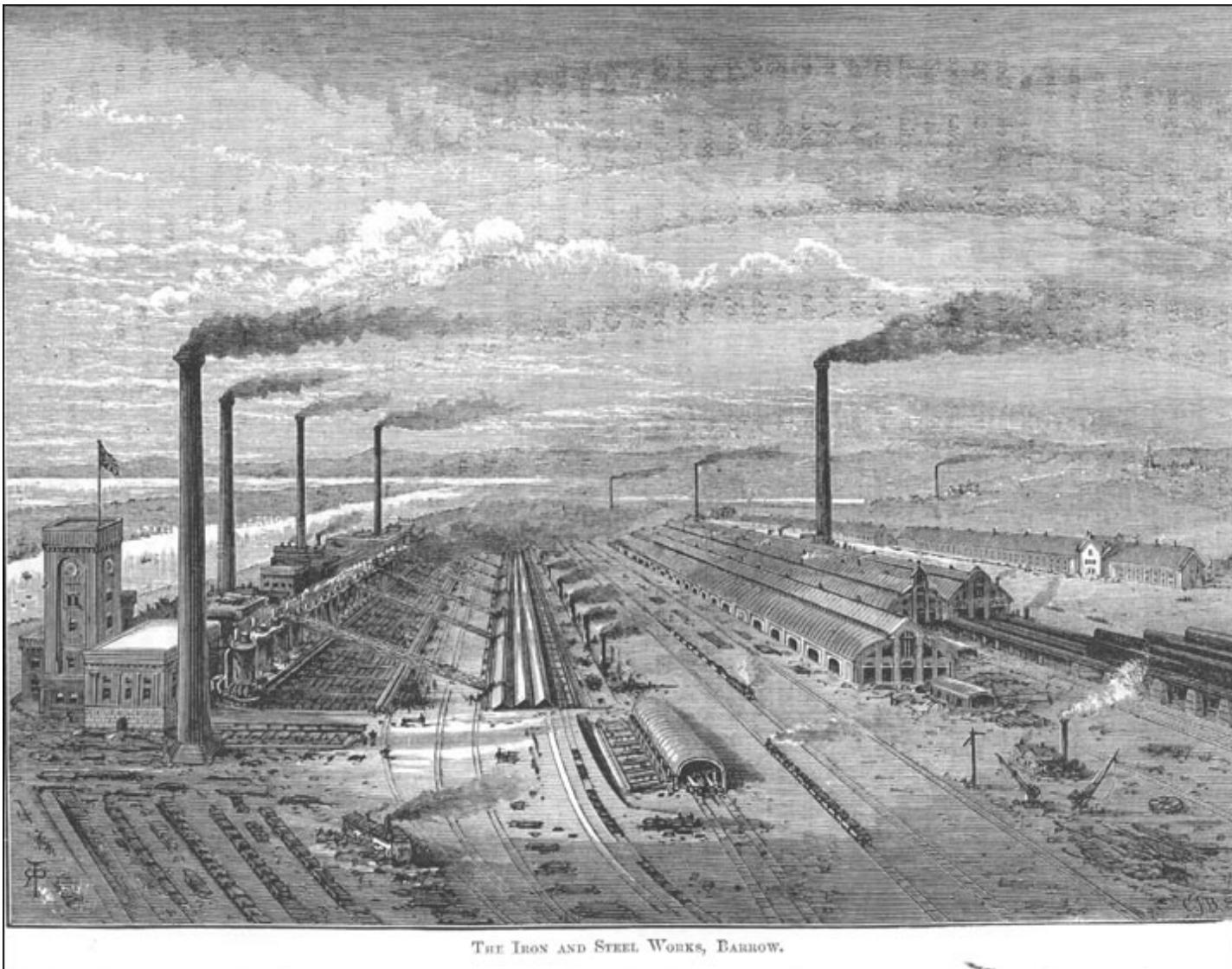
**Significance**  
**S**tudents learn more when they are actively engaged in the classroom than they do in a passive lecture environment. Extensive research supports this observation, especially in college-level science courses (1–6). Research also shows that active teaching strategies increase lecture attendance, engagement, and students' acquisition of expert attitudes toward the discipline (3, 7–9). Despite this overwhelming evidence, most instructors still use traditional methods, at least in large-enrollment college courses (10–12).

7:46 PM · Oct 22, 2020 · Twitter for iPhone

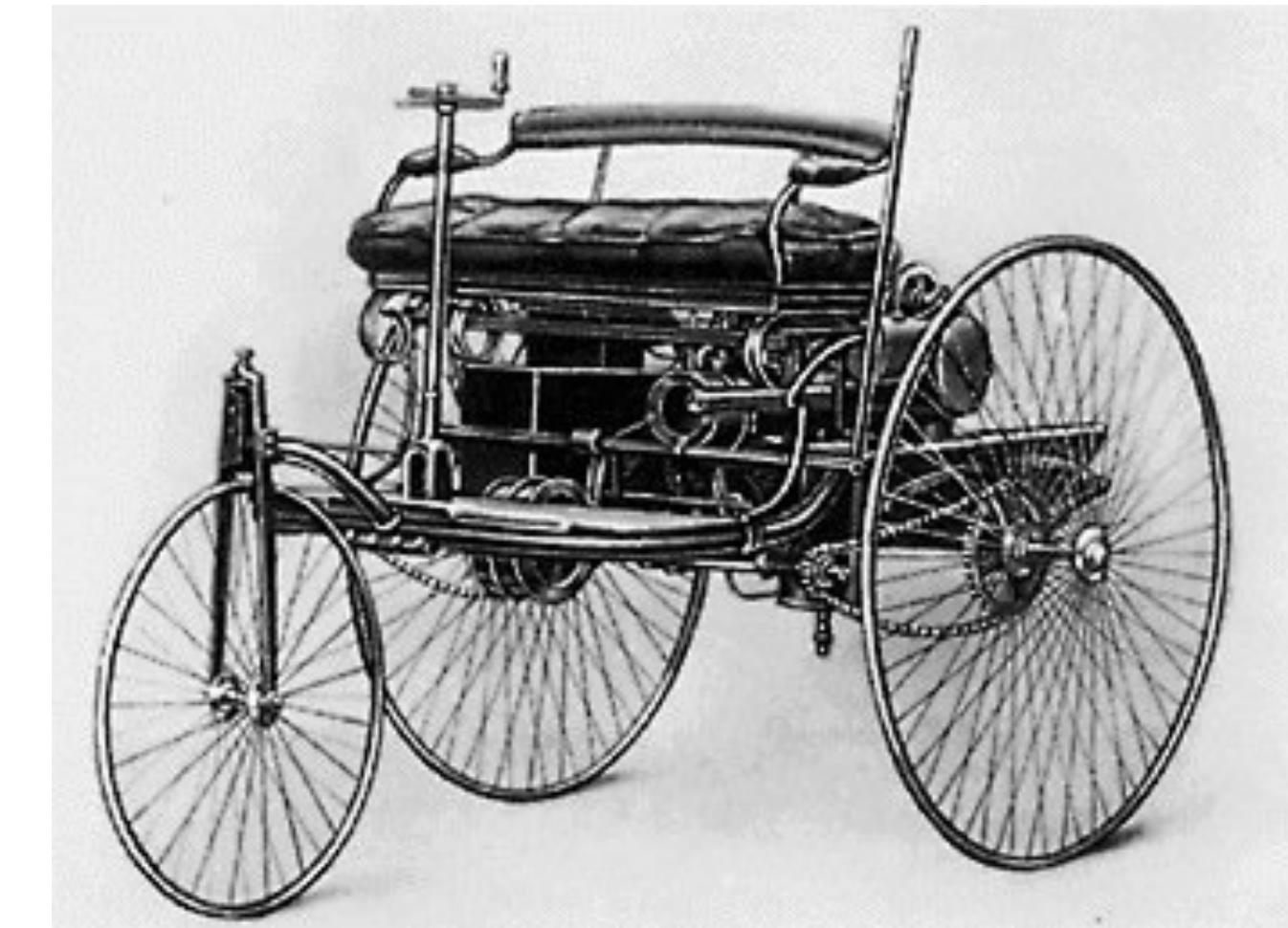
<https://twitter.com/emollick/status/1319424956241936384?s=20>

# Why visualization?

*1854, London*



*Industrial revolution  
(1760~1840)*

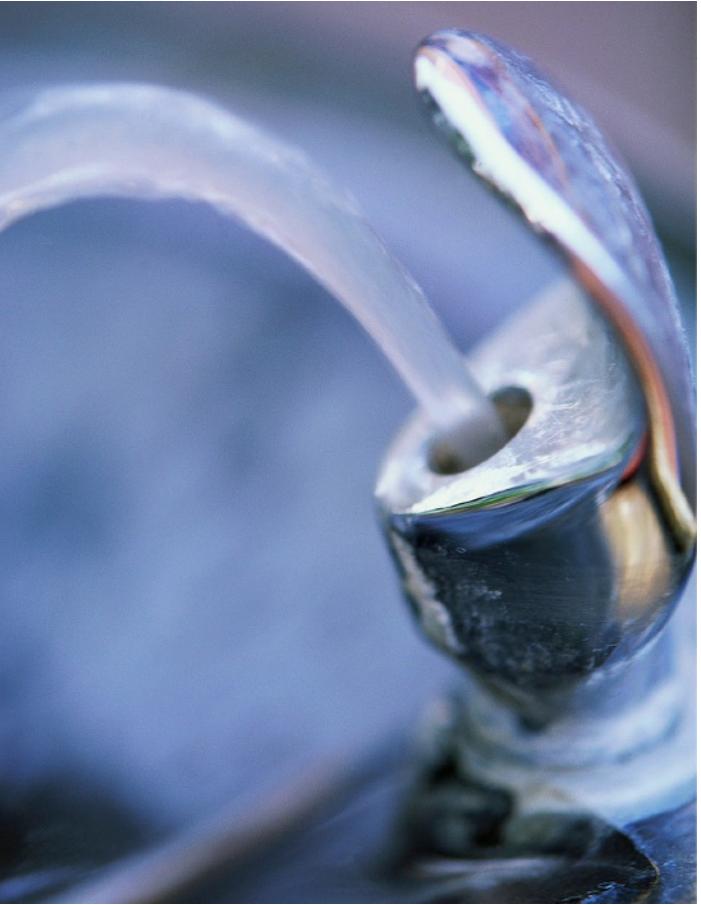


**Benz (1885)**

**1854, London**



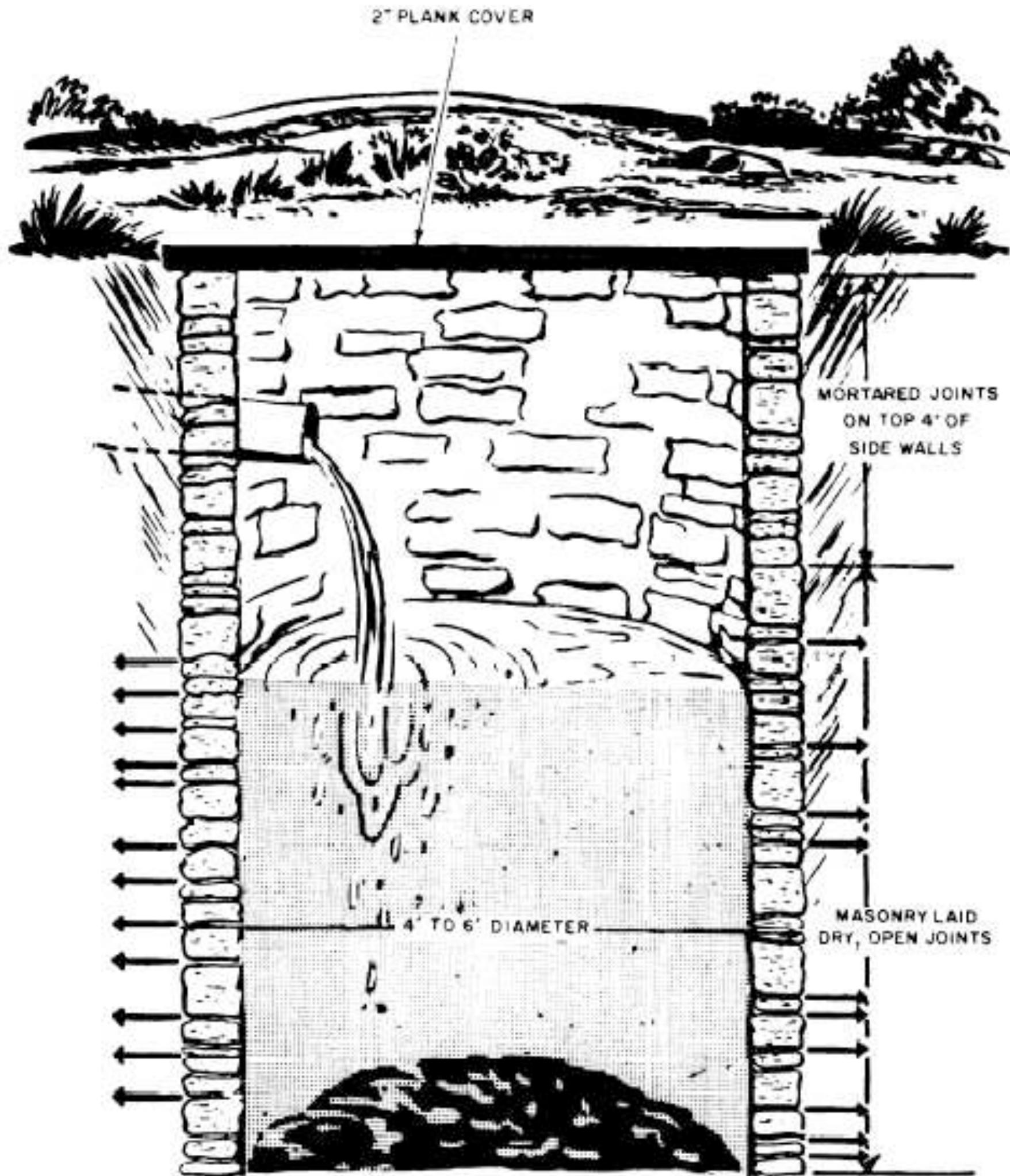
*Urbanization*



*but no  
infrastructure*

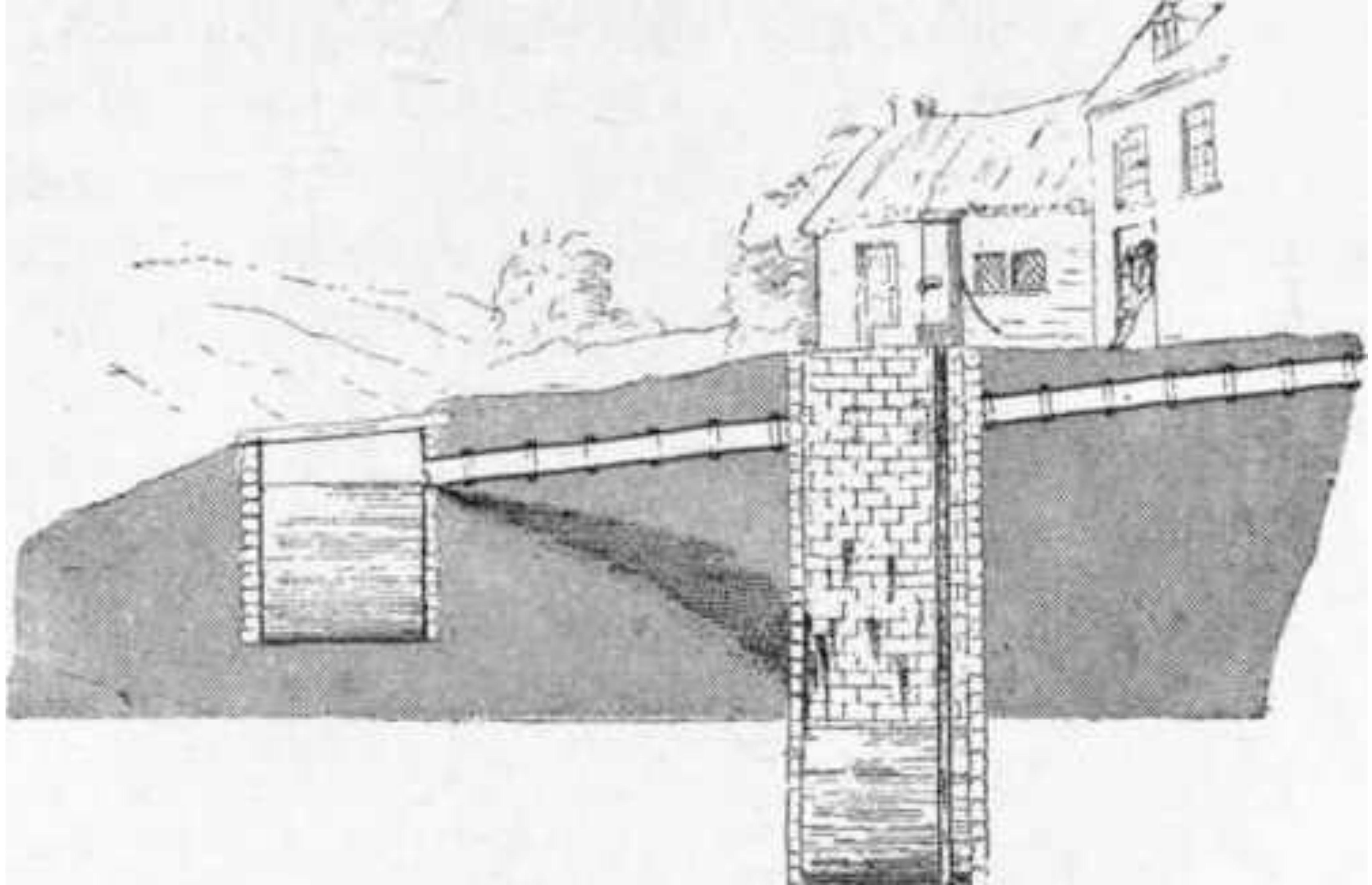
# *“Nightmen”*





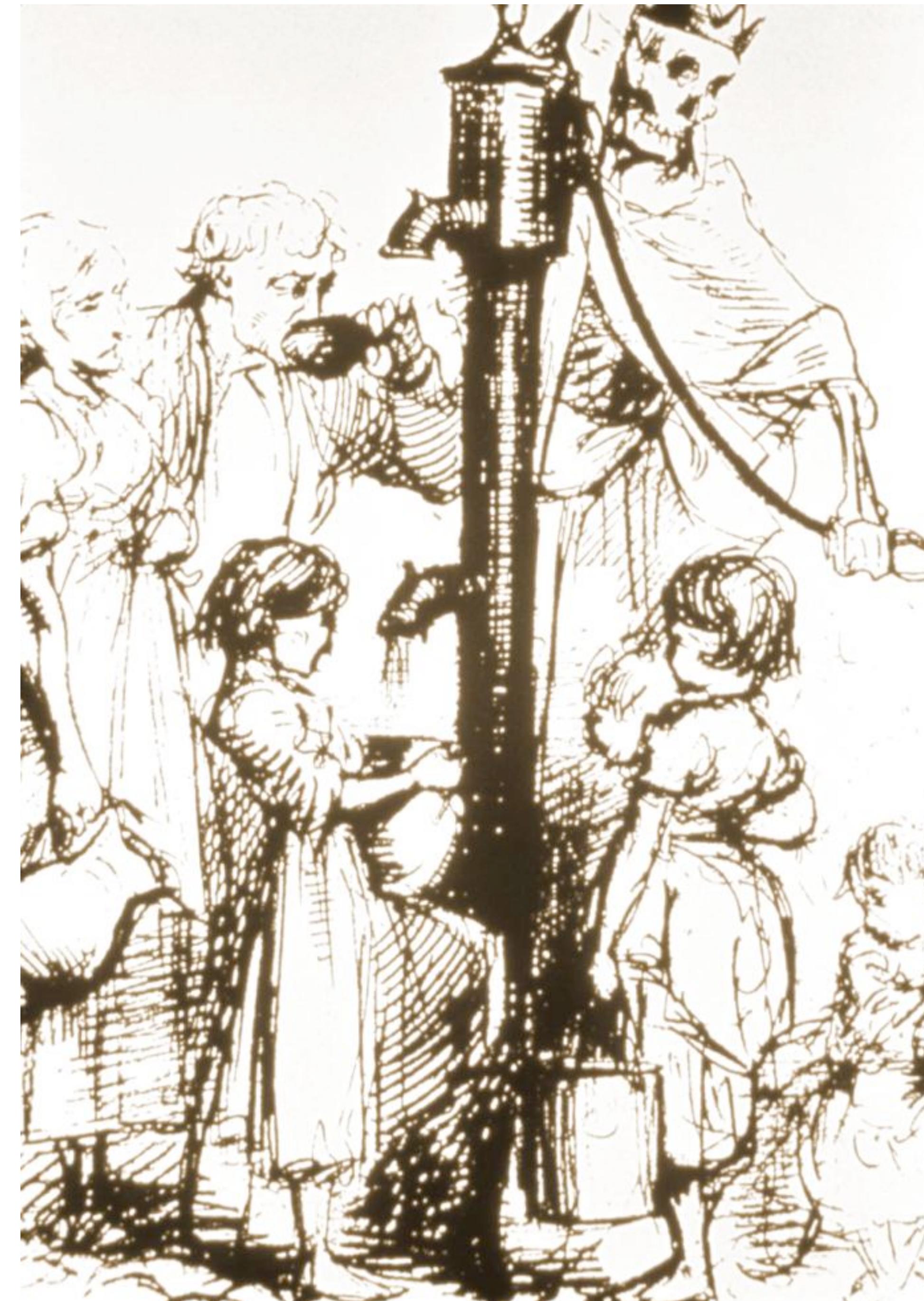
*“Night soil”*

*Cesspool*



*1854 London  
Cholera outbreak*

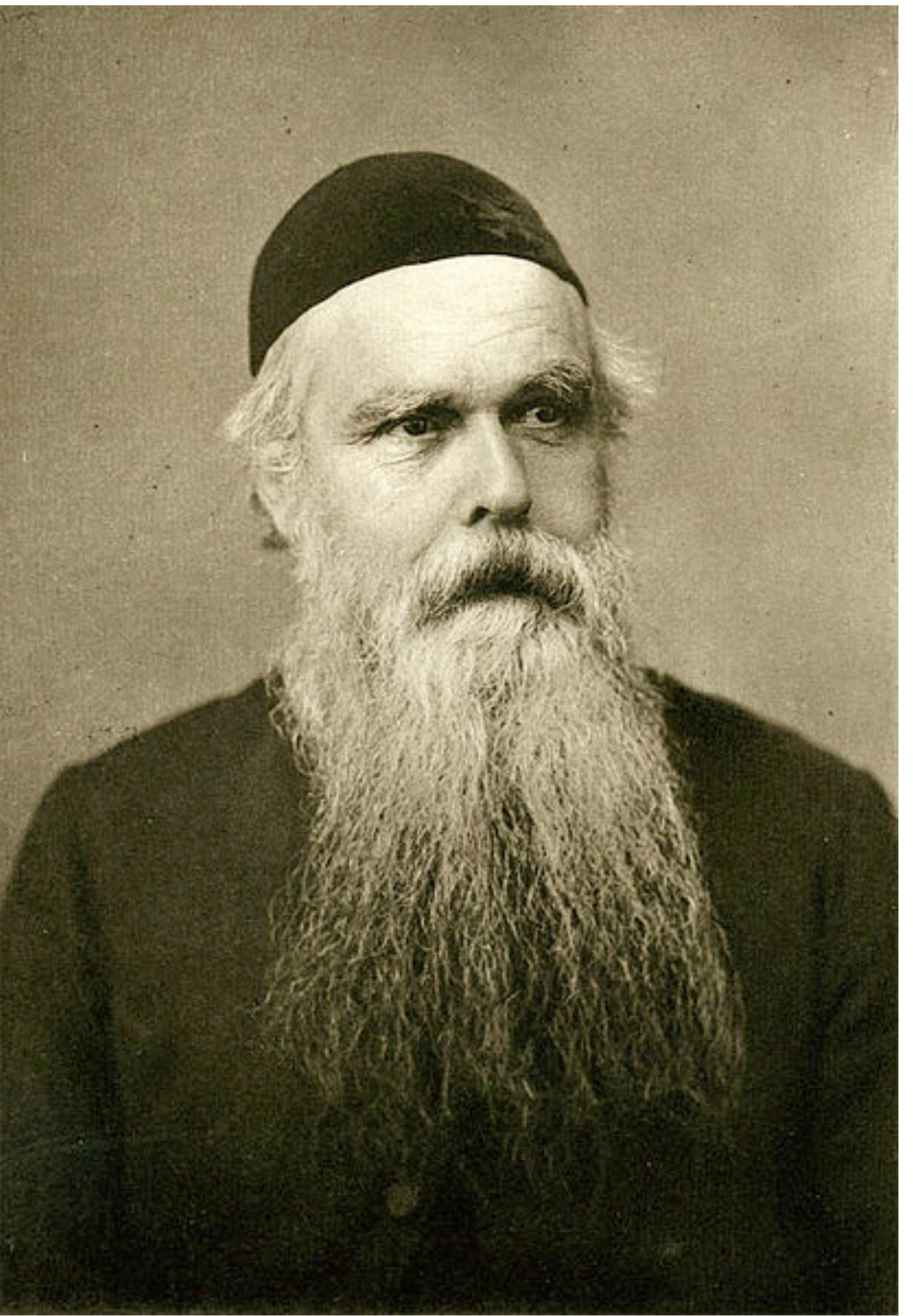
*127 died in three  
days,  
616 died at the end.*



“*Miasma (bad air) theory*”

*vs.*

“*Germ theory*”



# *Miasma!*

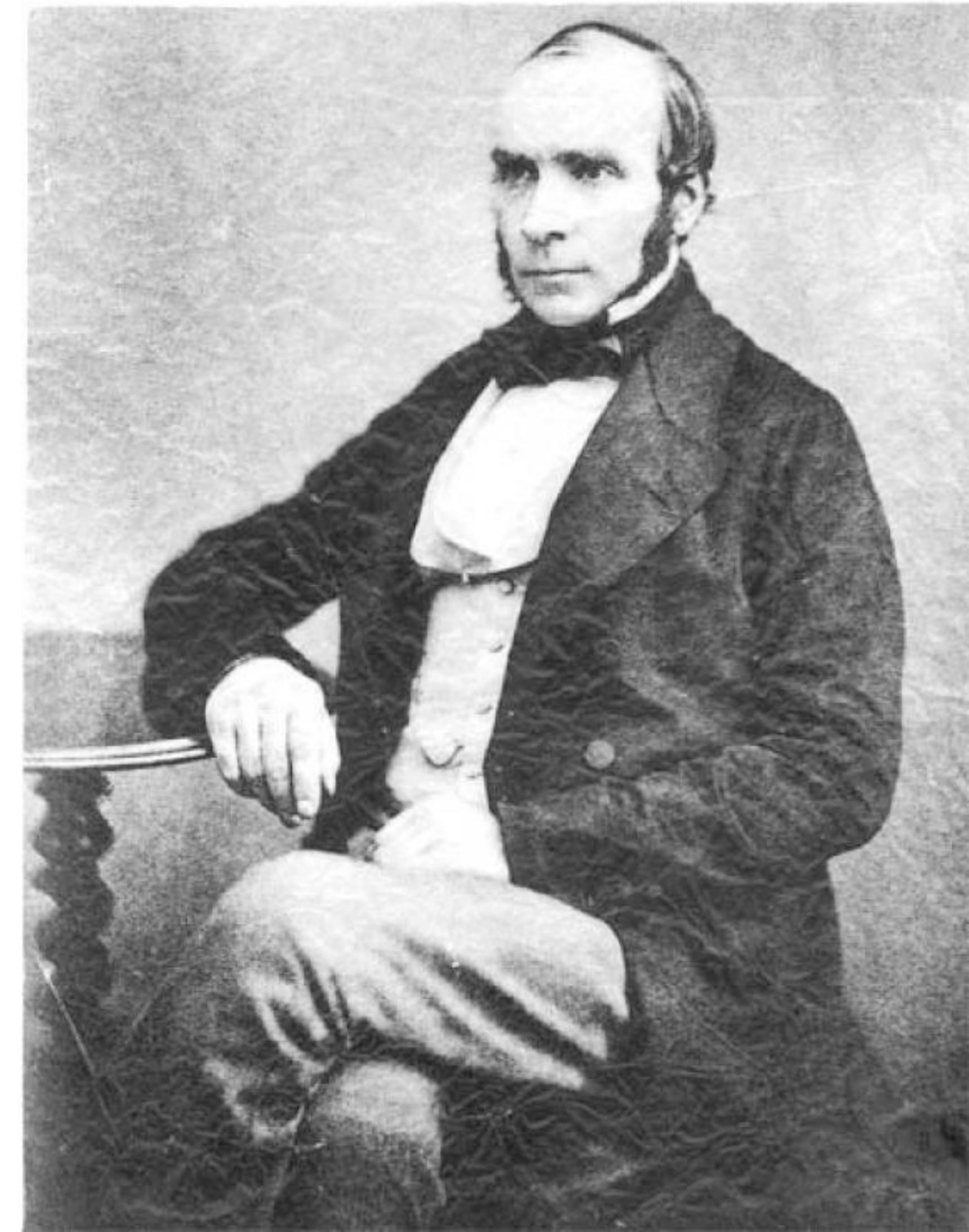
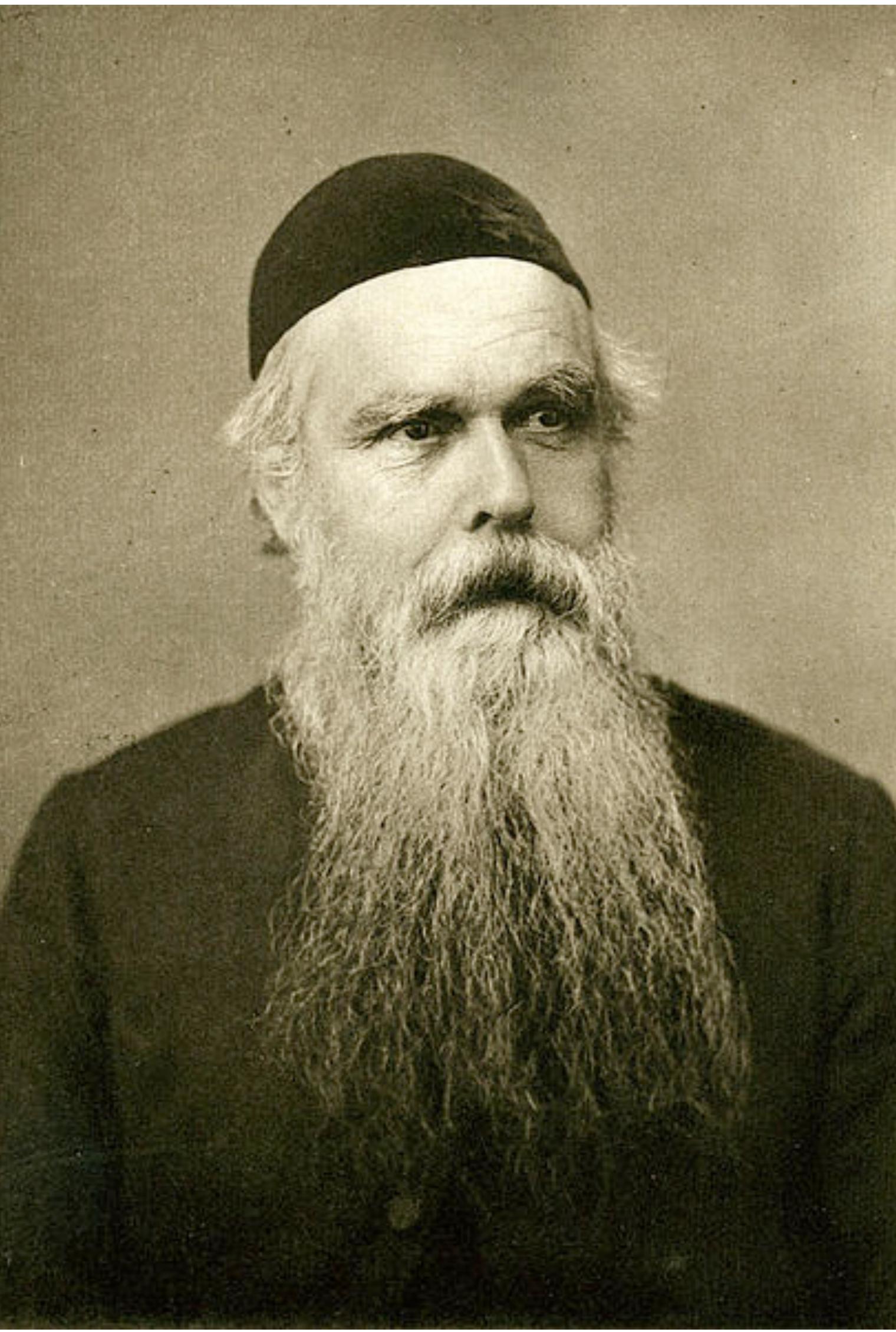
*Rev. Henry Whitehead*



*Germs!*

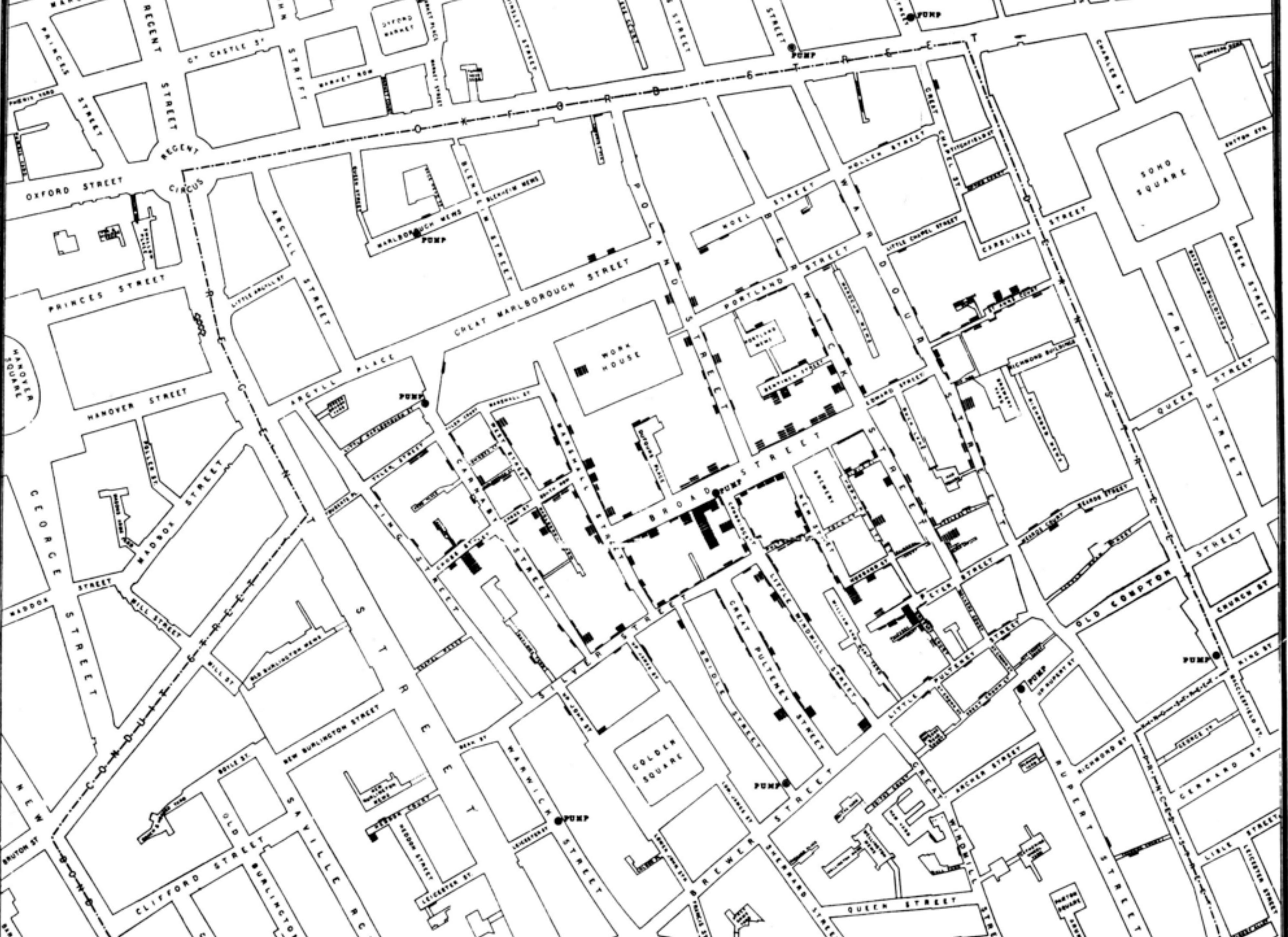
*John Snow*

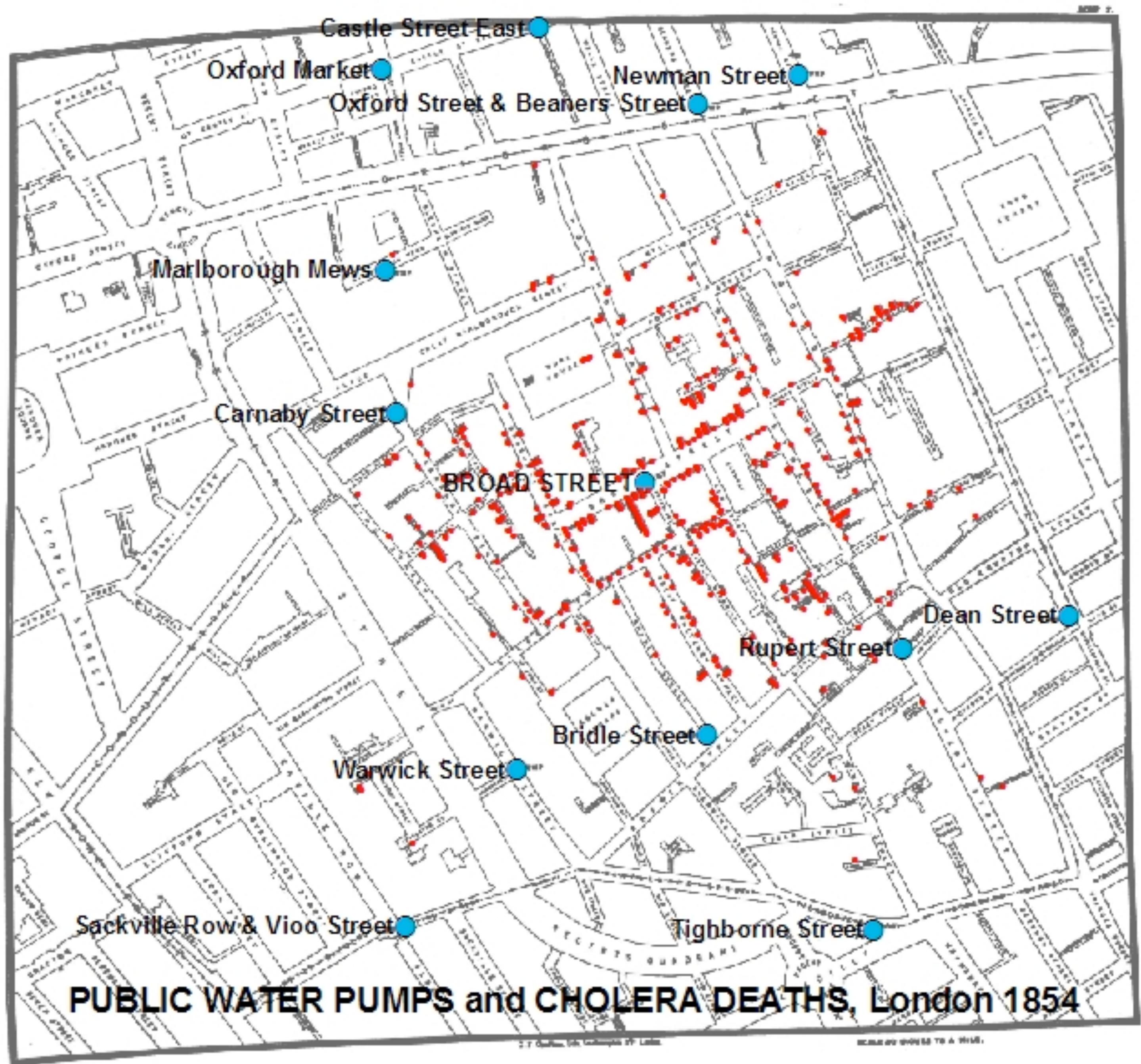




**What he (they) did  
changed the course of human race.**

**What did they do?**

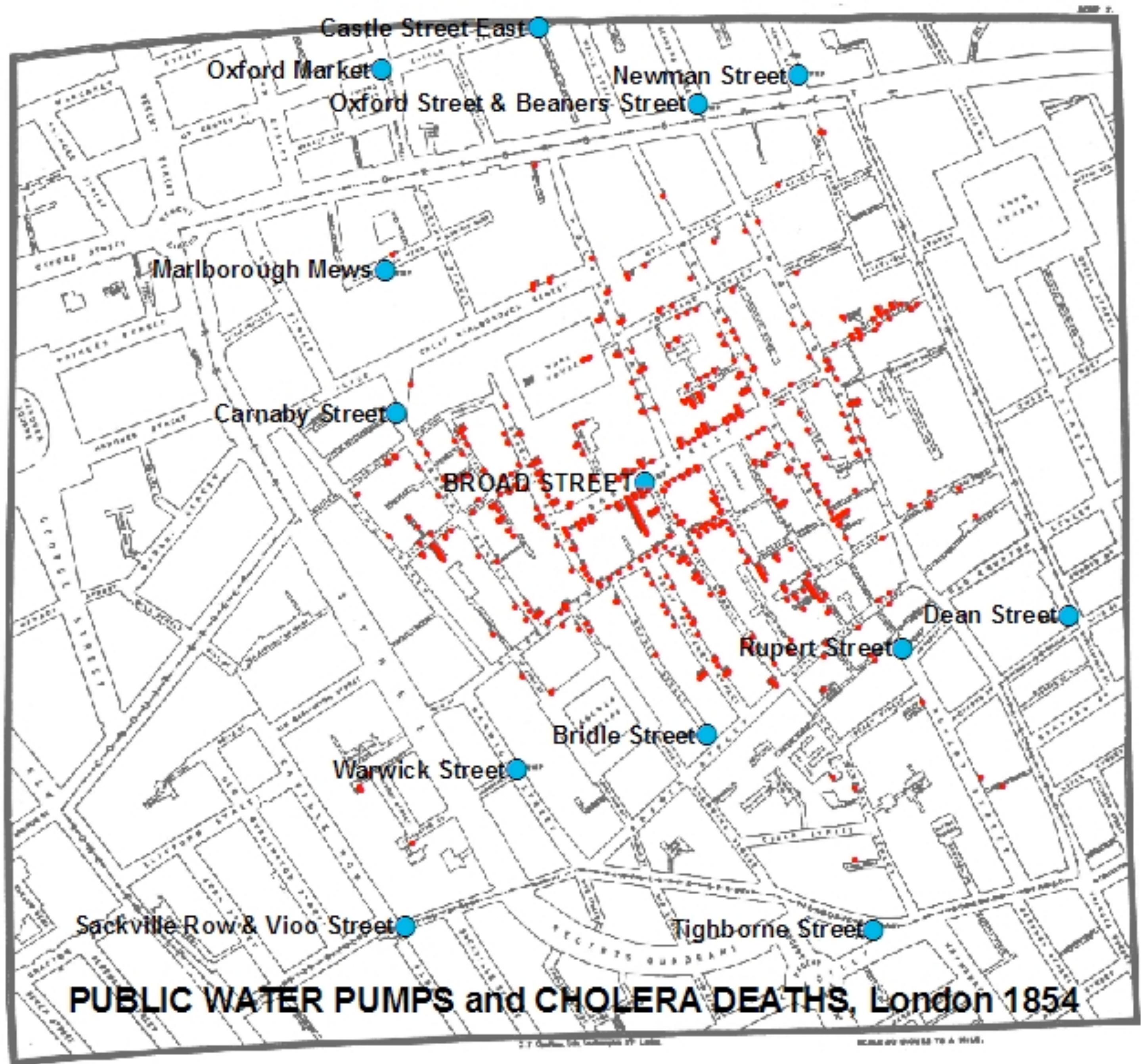


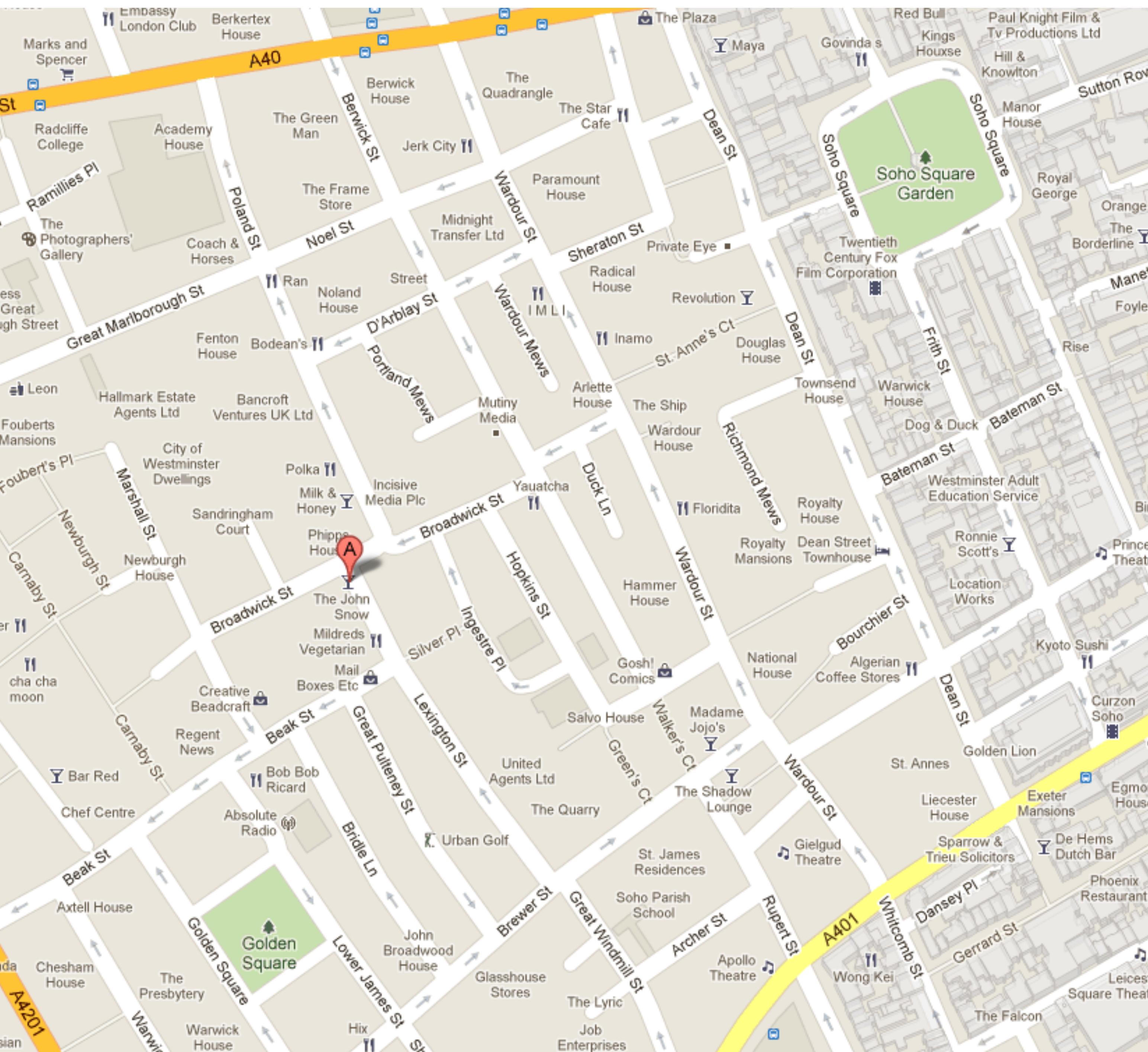


*On proceeding to the spot, I found that nearly all the deaths had taken place within a short distance of the [Broad Street] pump. There were only ten deaths in houses situated decidedly nearer to another street-pump. In five of these cases the families of the deceased persons informed me that they always sent to the pump in Broad Street, as they preferred the water to that of the pumps which were nearer. In three other cases, the deceased were children who went to school near the pump in Broad Street...*

...

—John Snow, letter to the editor of the *Medical Times and Gazette*







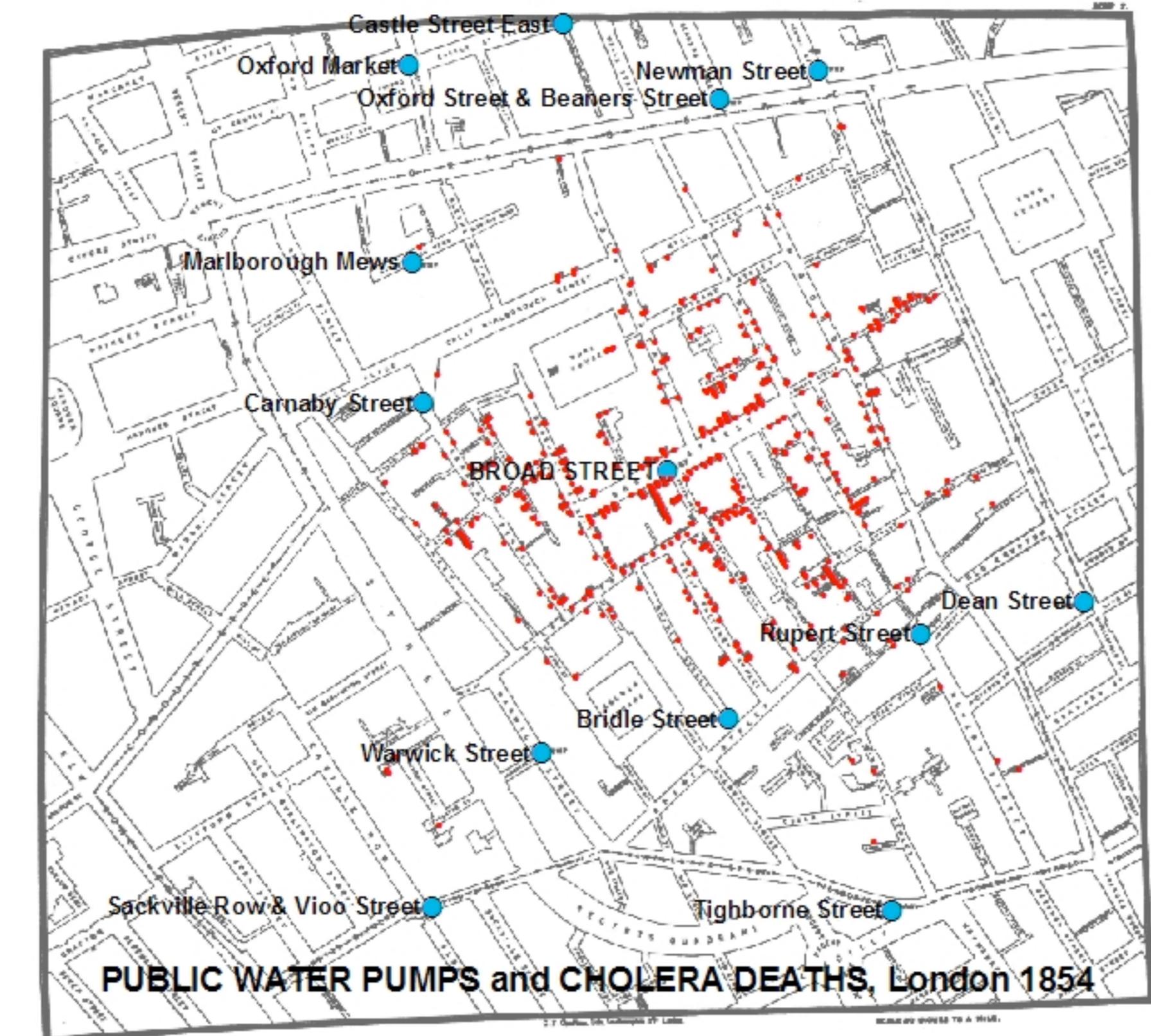


**Now imagine...**

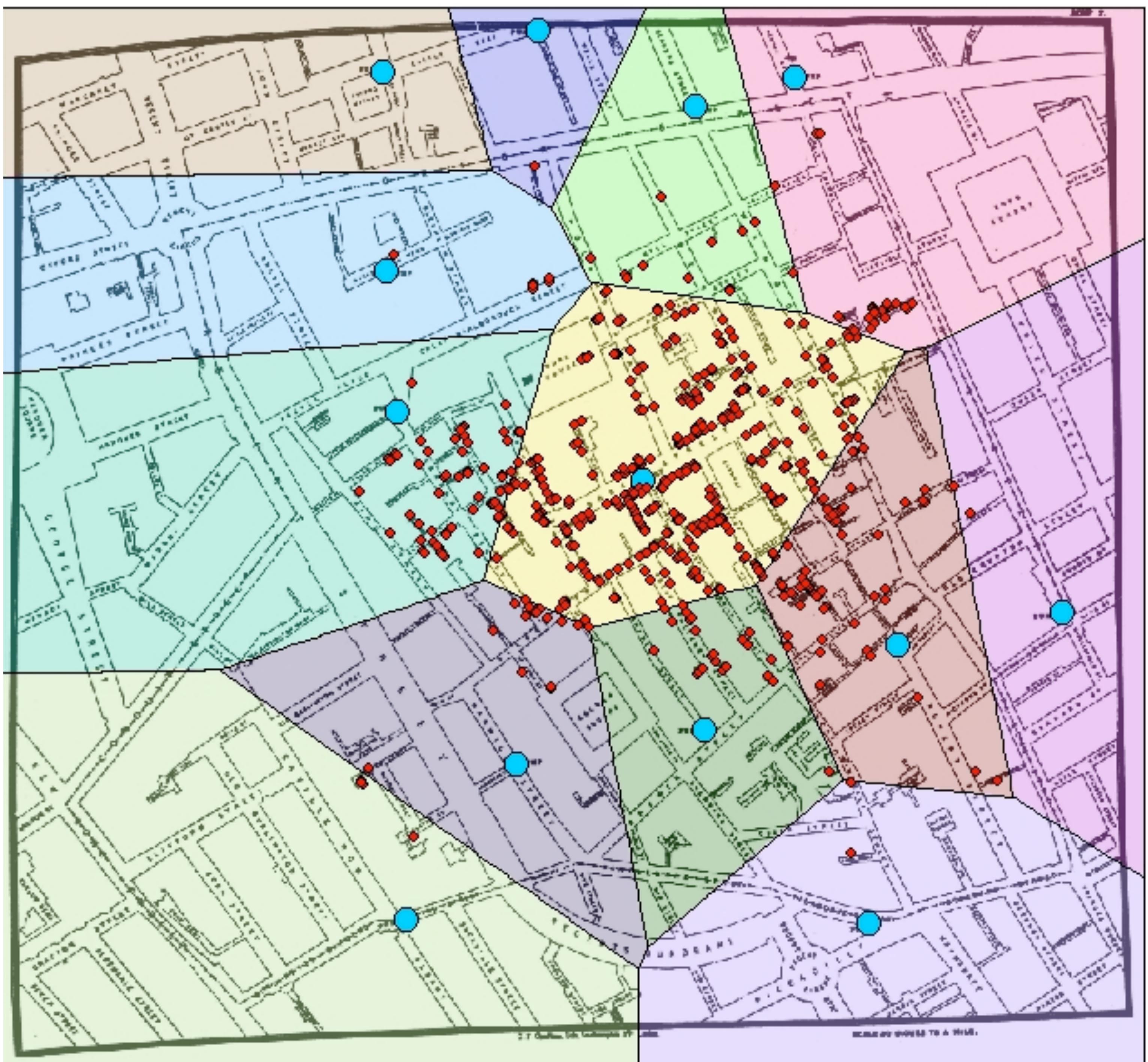
**Can you make  
Dr. Snow's job easier?**

# How do you draw a map to reveal

“The nearest pump  
for any given  
location”?







# “Voronoi diagram”

<http://bl.ocks.org/mbostock/4060366>

Can't we just use **numbers** and  
**statistics**?

I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

### “Anscombe's quartet”

N = 11

mean of X's = 9.0

mean of Y's = 7.5

equation of regression line:  $Y = 3 + 0.5X$

standard error of estimate of slope = 0.118

t = 4.24

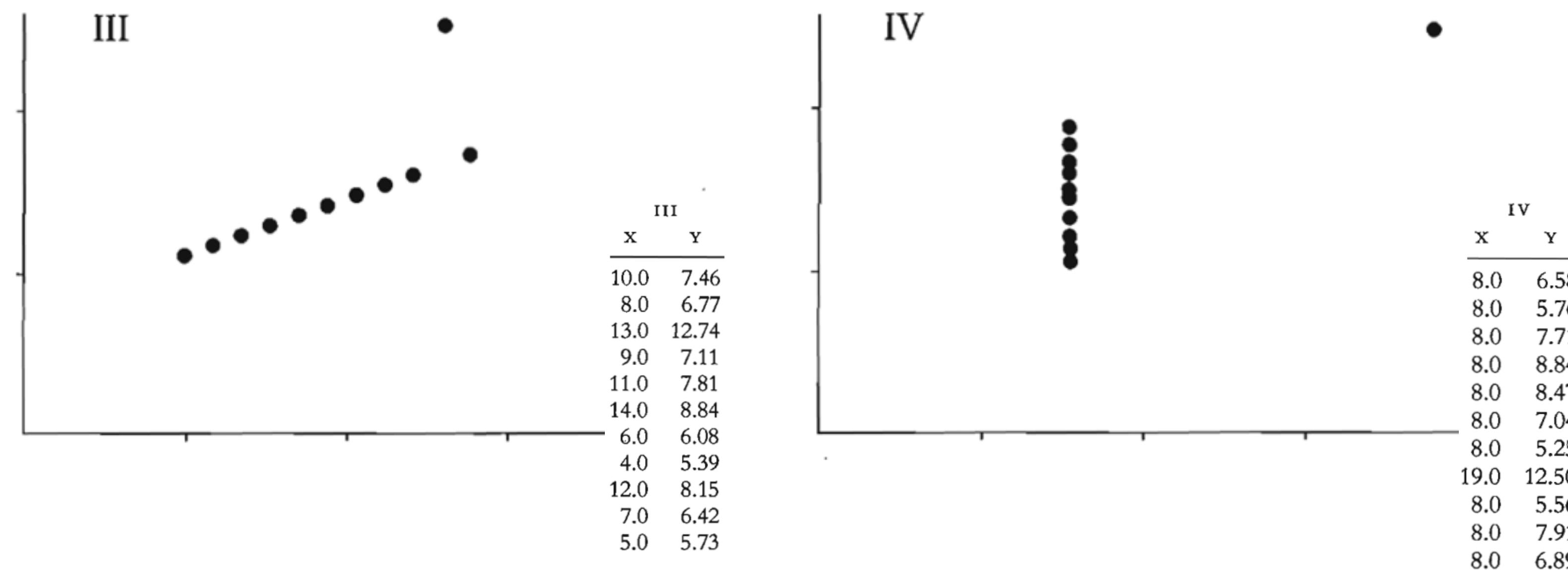
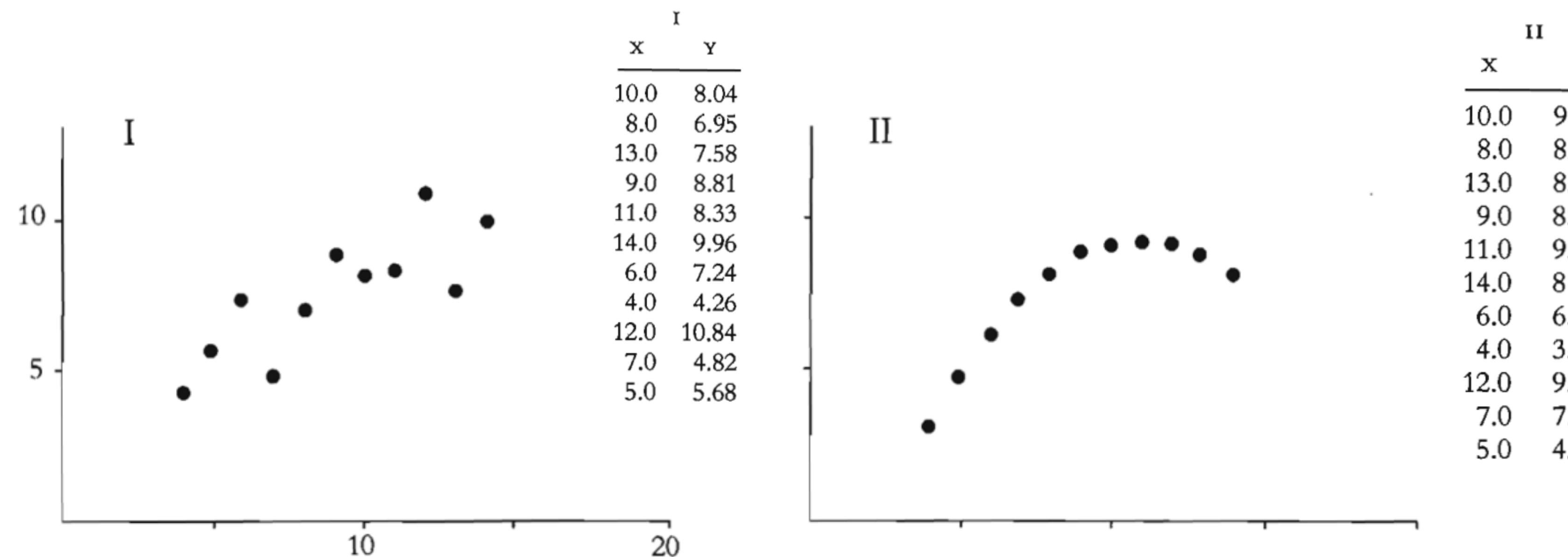
sum of squares  $\sum (X - \bar{X})^2 = 110.0$

regression sum of squares = 27.50

residual sum of squares of Y = 13.75

correlation coefficient = .82

$r^2 = .67$

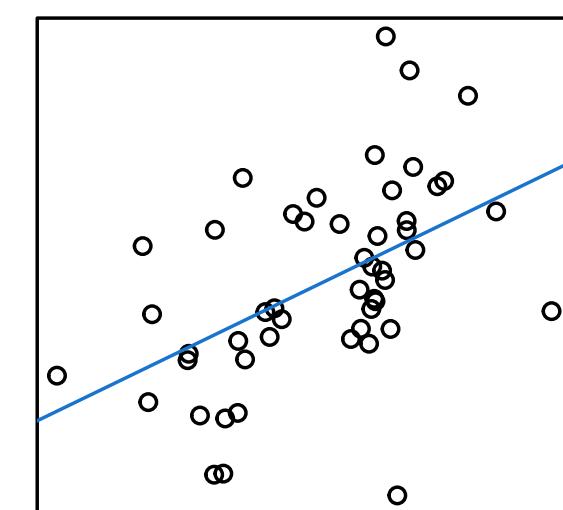


“X and Y are correlated with Pearson’s correlation coefficient of 0.5.”

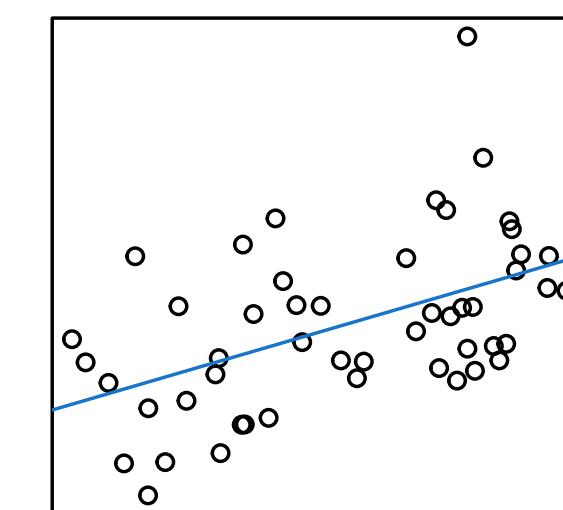
Can you sketch how this would look like?

## All correlations: $r(50) = 0.5$

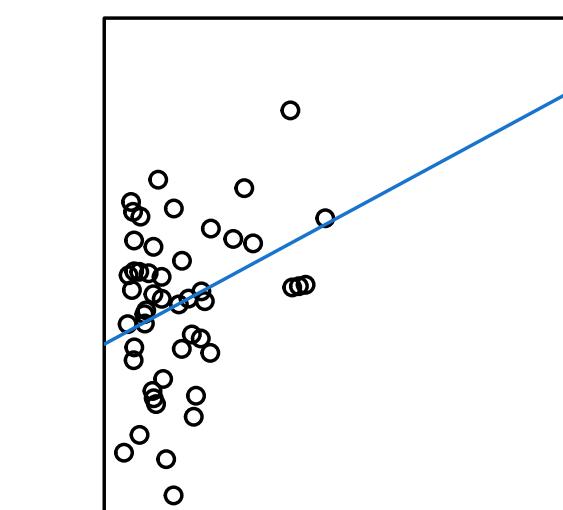
(1) Normal x, normal residuals



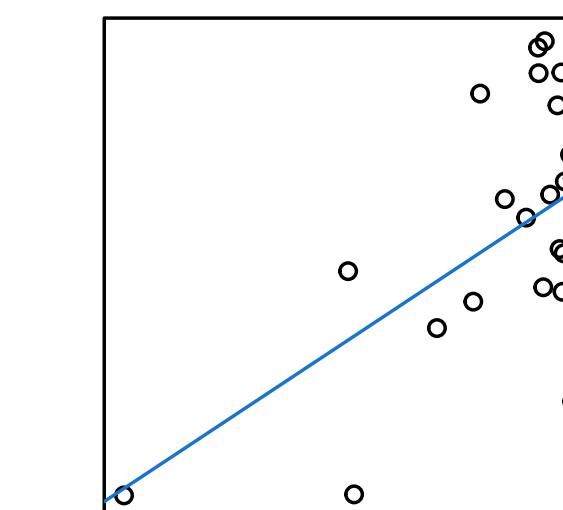
(2) Uniform x, normal residuals



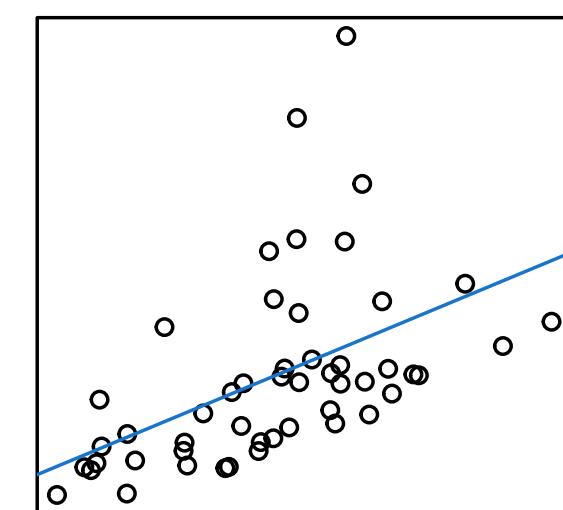
(3) +-skewed x, normal residuals



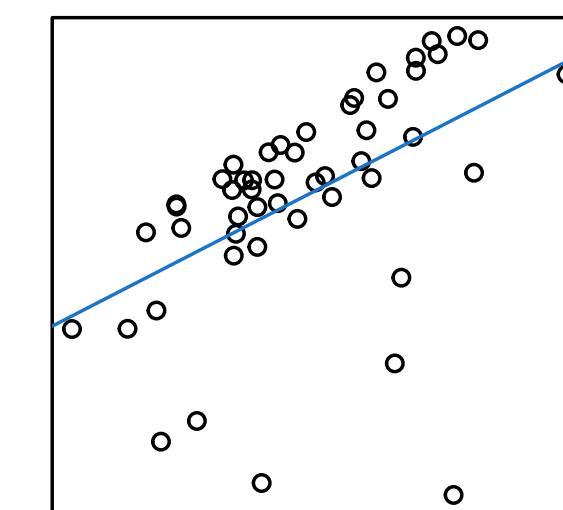
(4) --skewed x, normal residuals



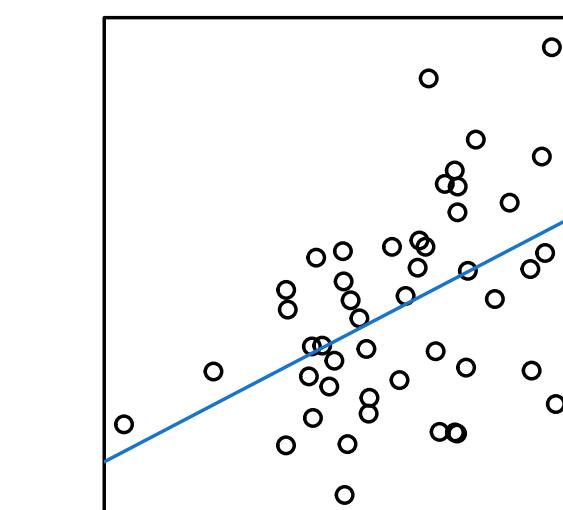
(5) Normal x, +-skewed residuals



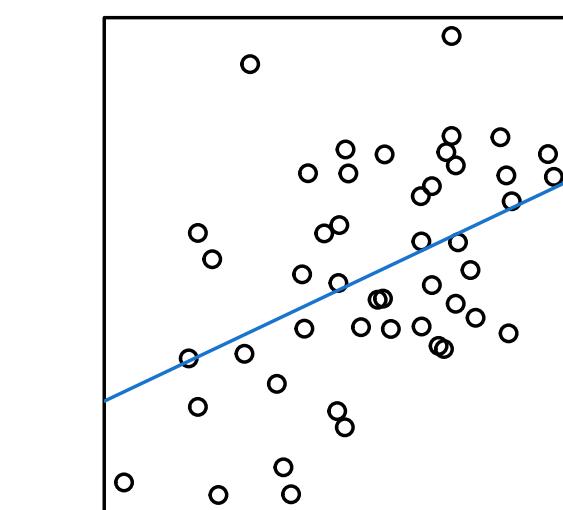
(6) Normal x, --skewed residuals



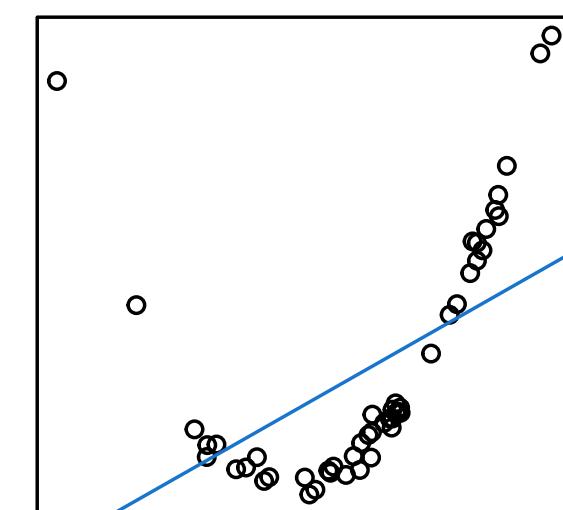
(7) Increasing spread



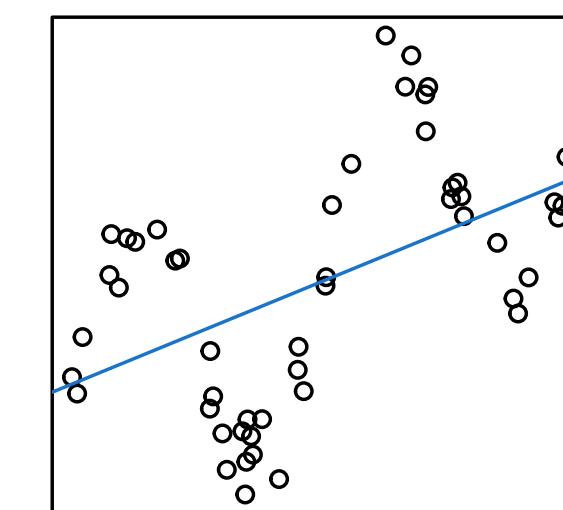
(8) Decreasing spread



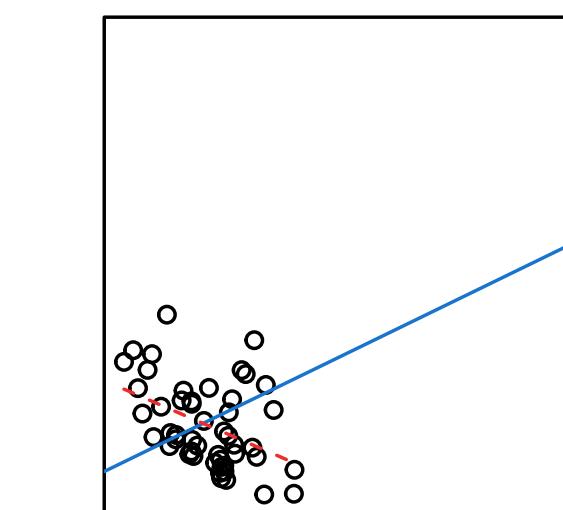
(9) Quadratic trend



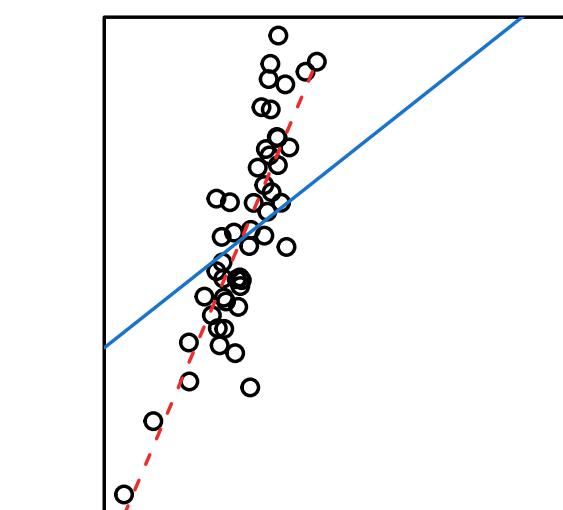
(10) Sinusoid relationship



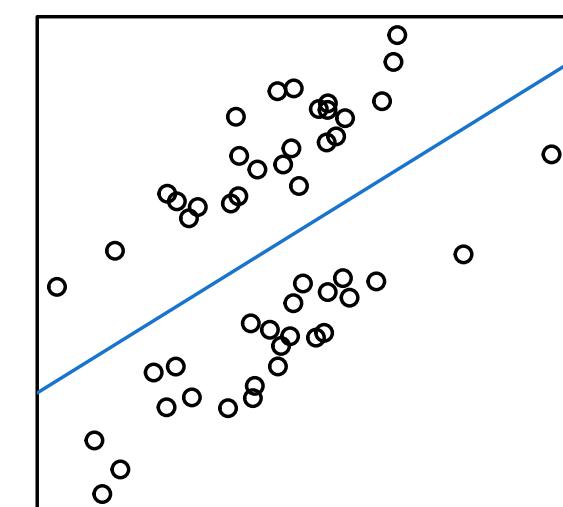
(11) A single positive outlier



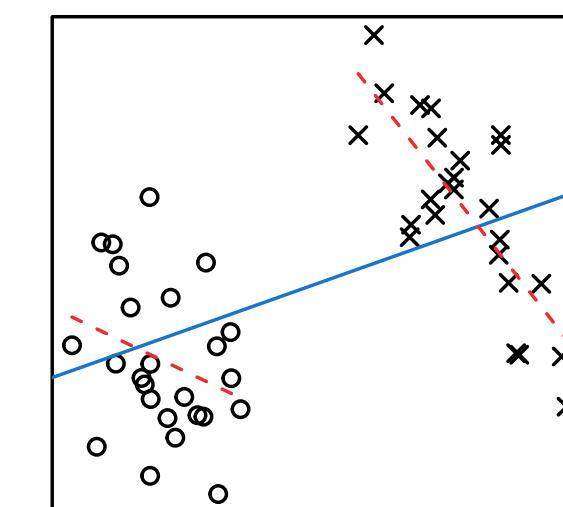
(12) A single negative outlier



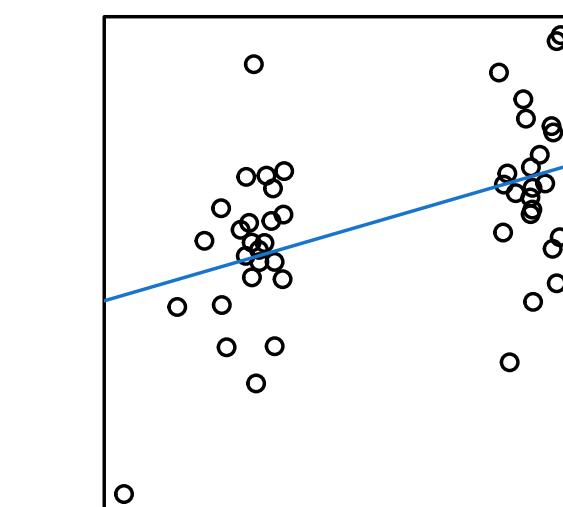
(13) Bimodal residuals



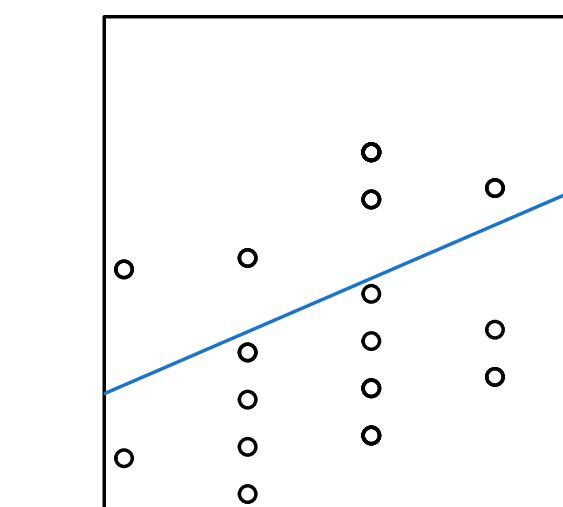
(14) Two groups

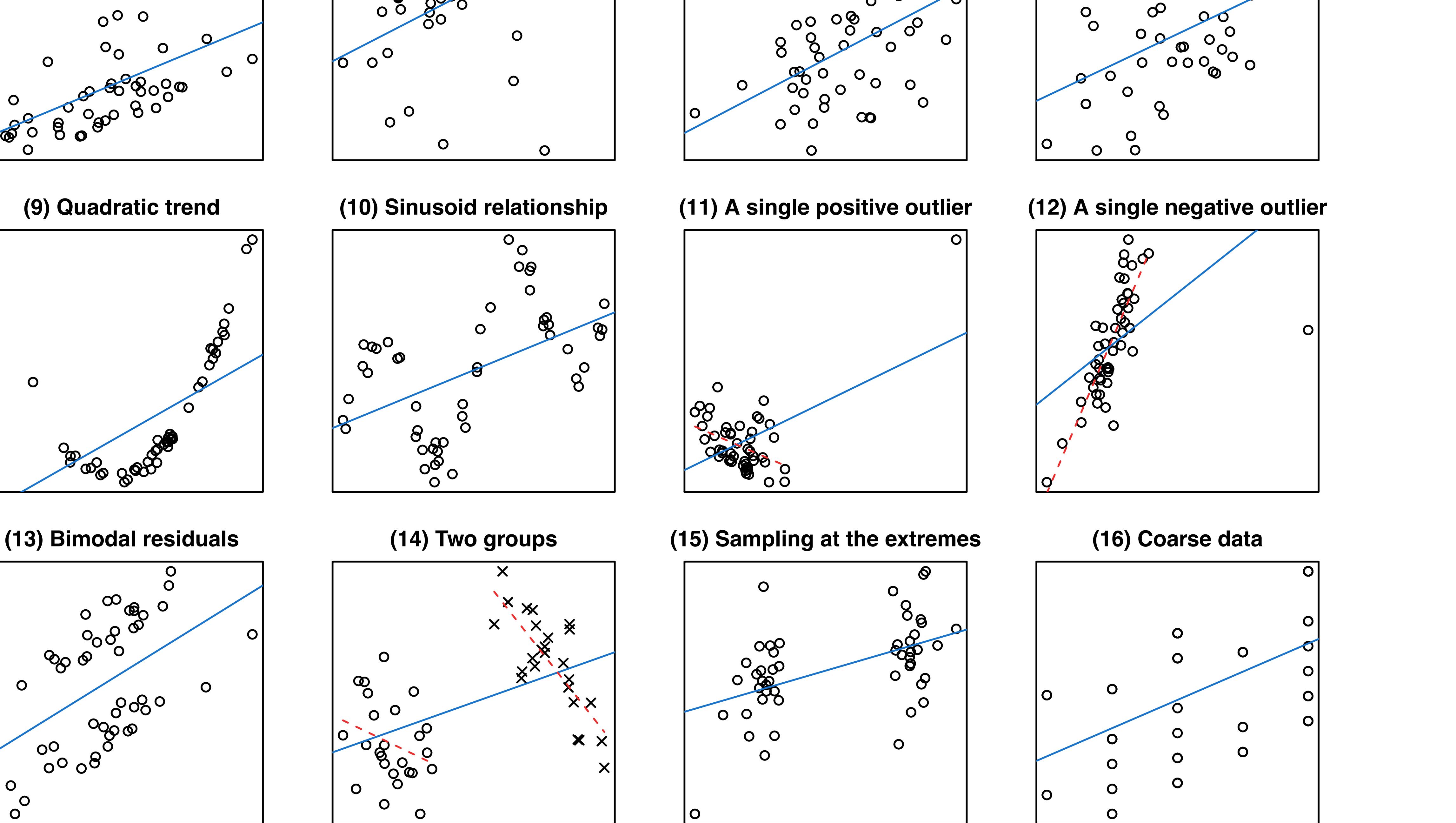


(15) Sampling at the extremes



(16) Coarse data



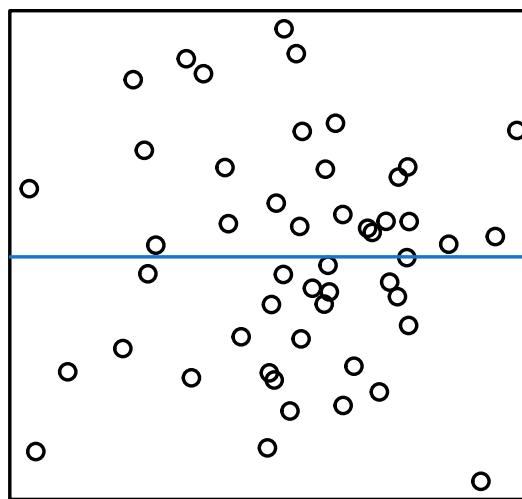


**How about zero correlation?**

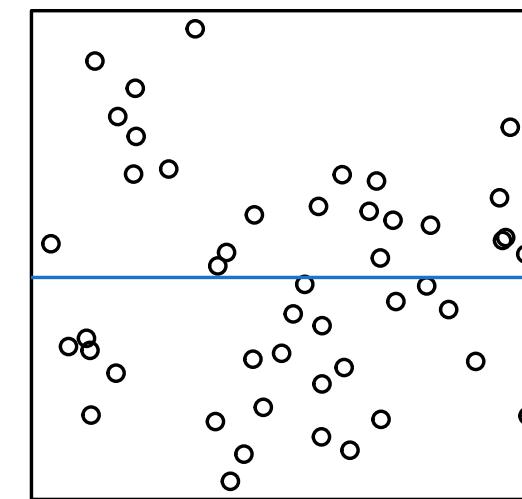
**Sketch?**

## All correlations: $r(50) = 0$

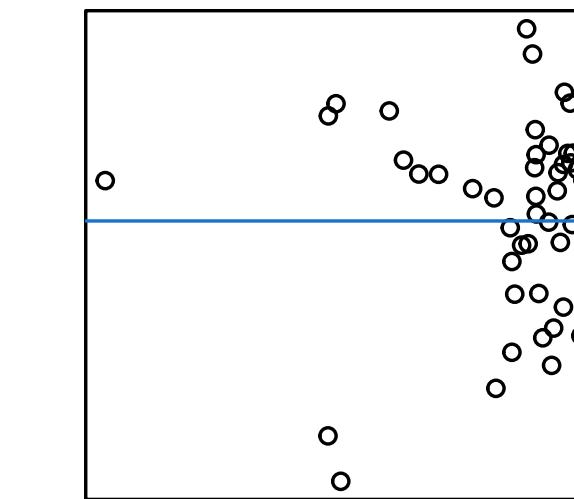
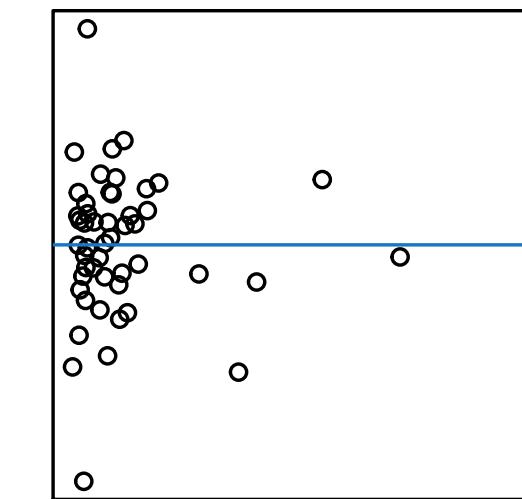
(1) Normal x, normal residuals



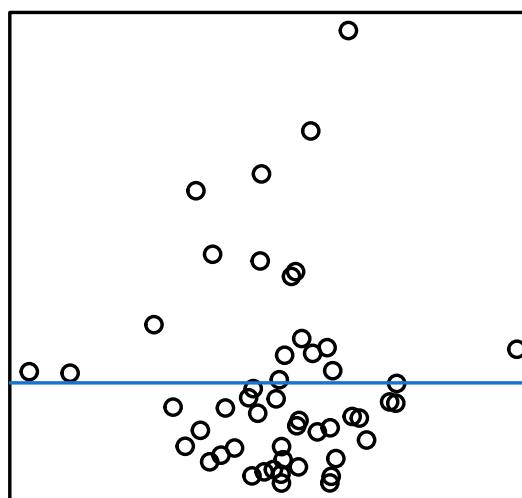
(2) Uniform x, normal residuals



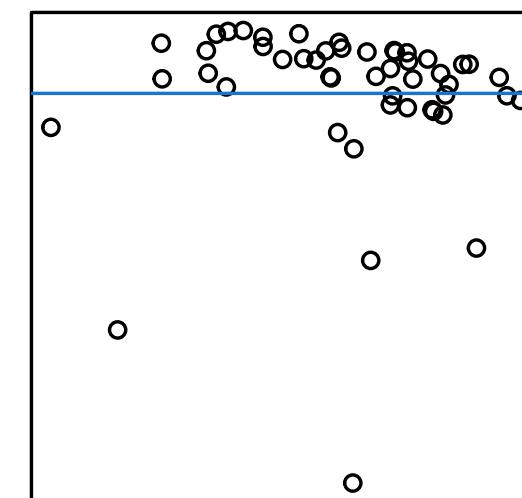
(3) +-skewed x, normal residuals (4) --skewed x, normal residuals



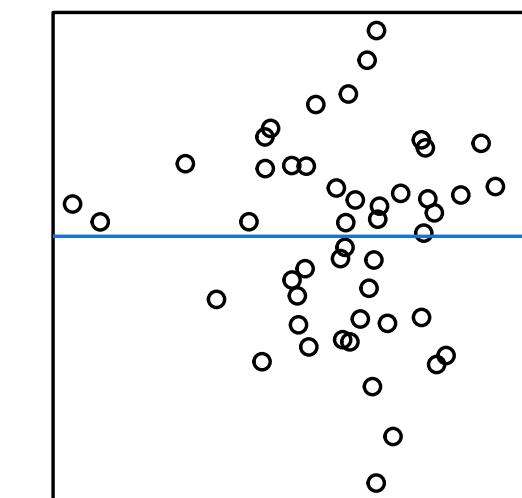
(5) Normal x, +-skewed residuals



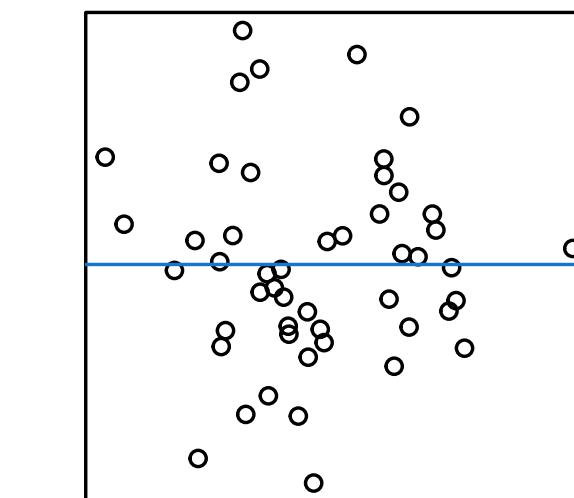
(6) Normal x, --skewed residuals



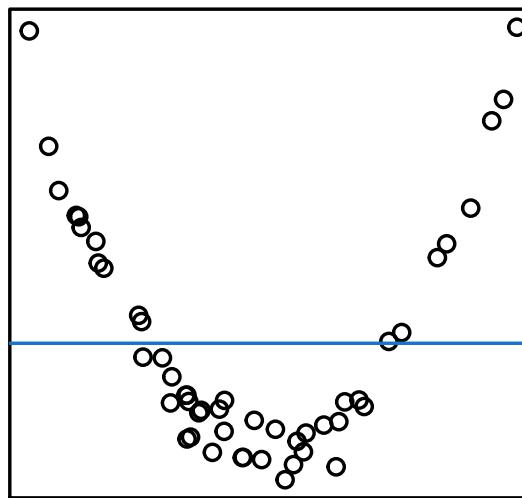
(7) Increasing spread



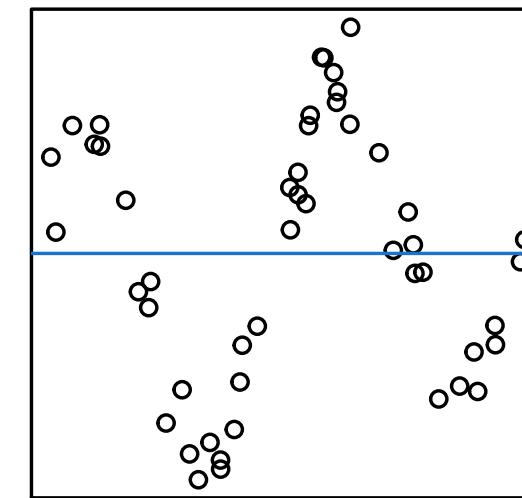
(8) Decreasing spread



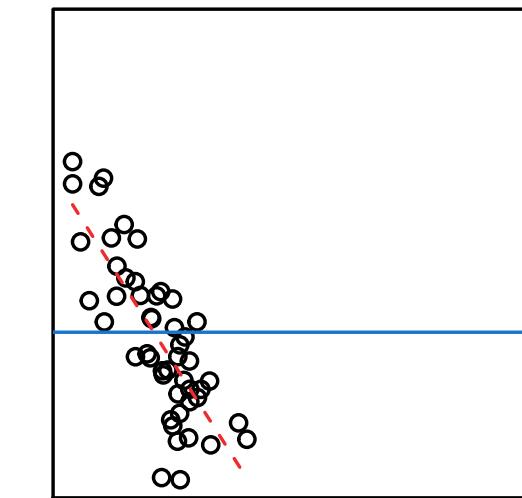
(9) Quadratic trend



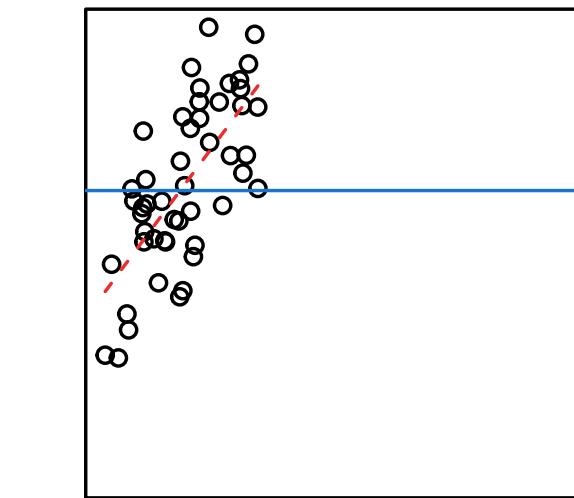
(10) Sinusoid relationship



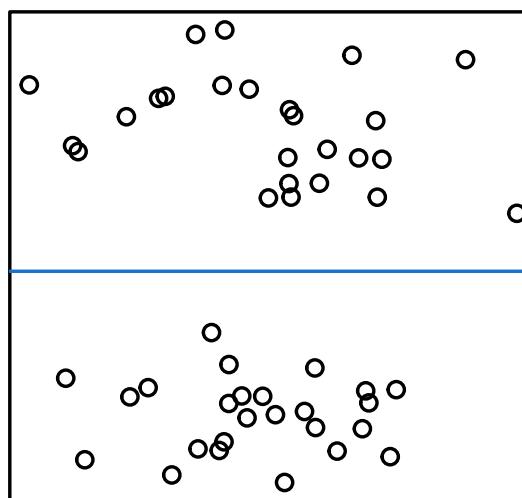
(11) A single positive outlier



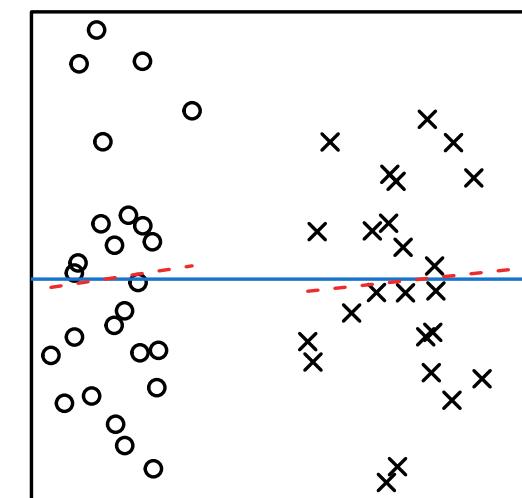
(12) A single negative outlier



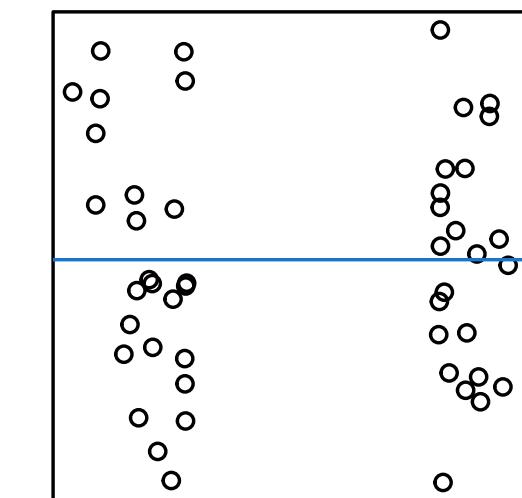
(13) Bimodal residuals



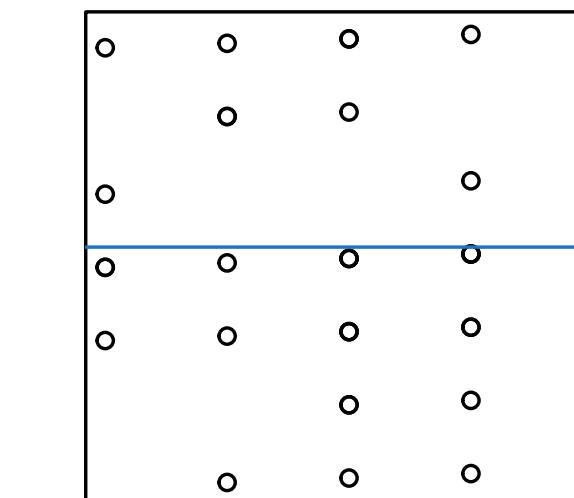
(14) Two groups

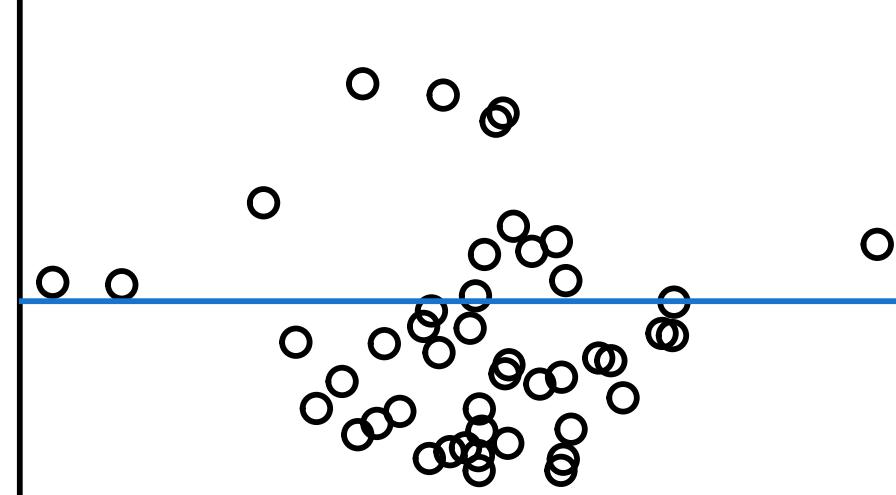


(15) Sampling at the extremes

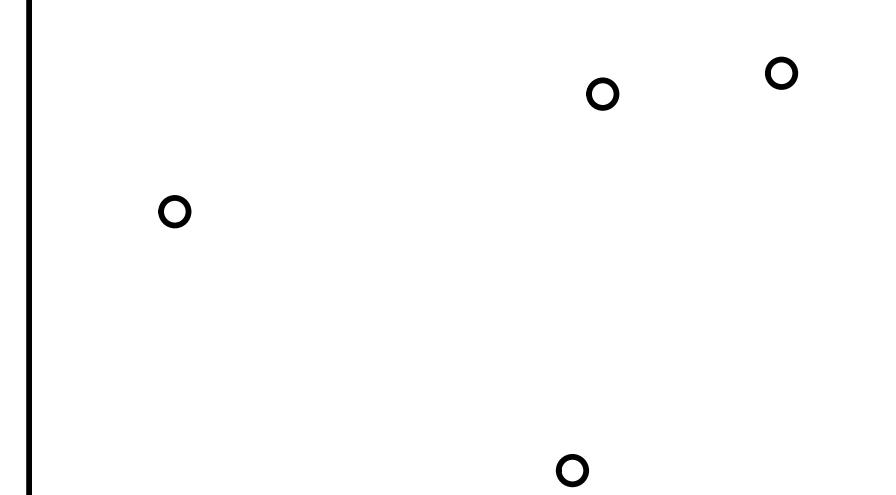


(16) Coarse data

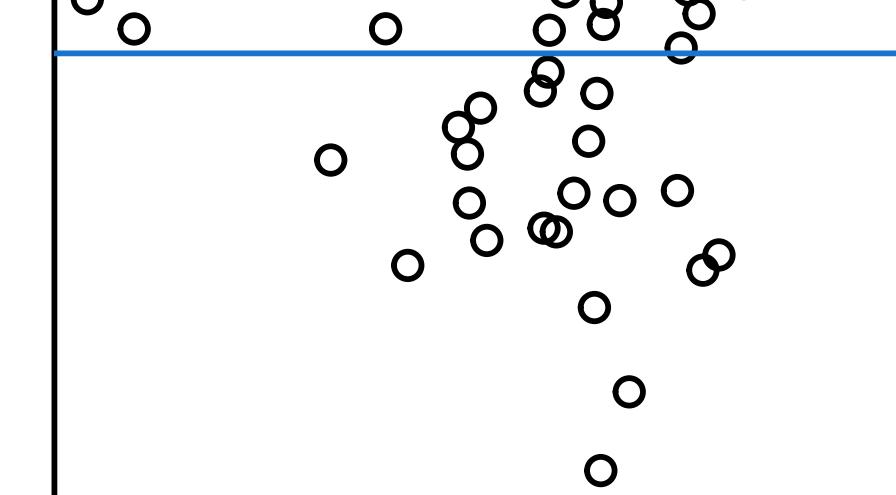




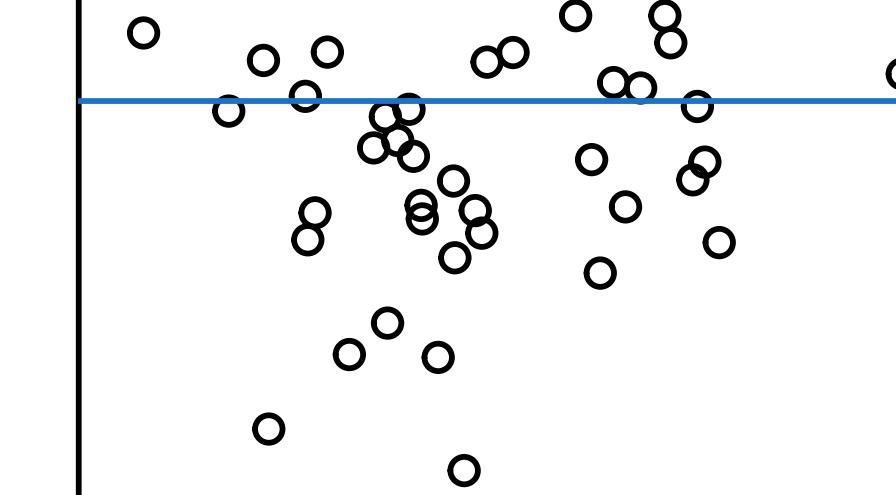
(9) Quadratic trend



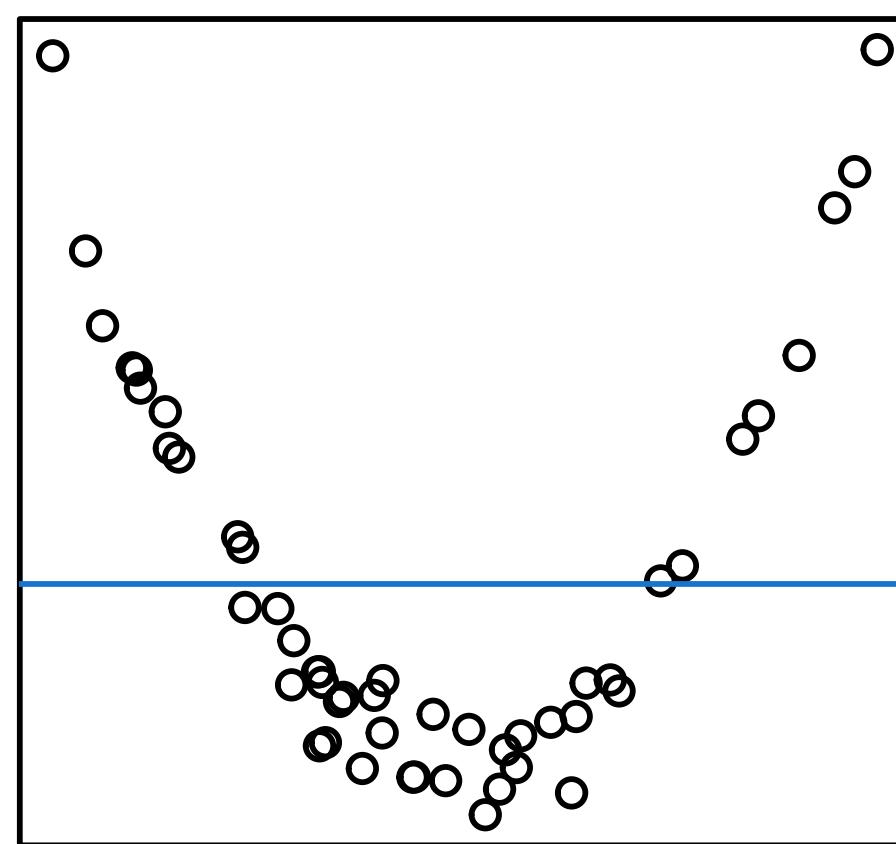
(10) Sinusoid relationship



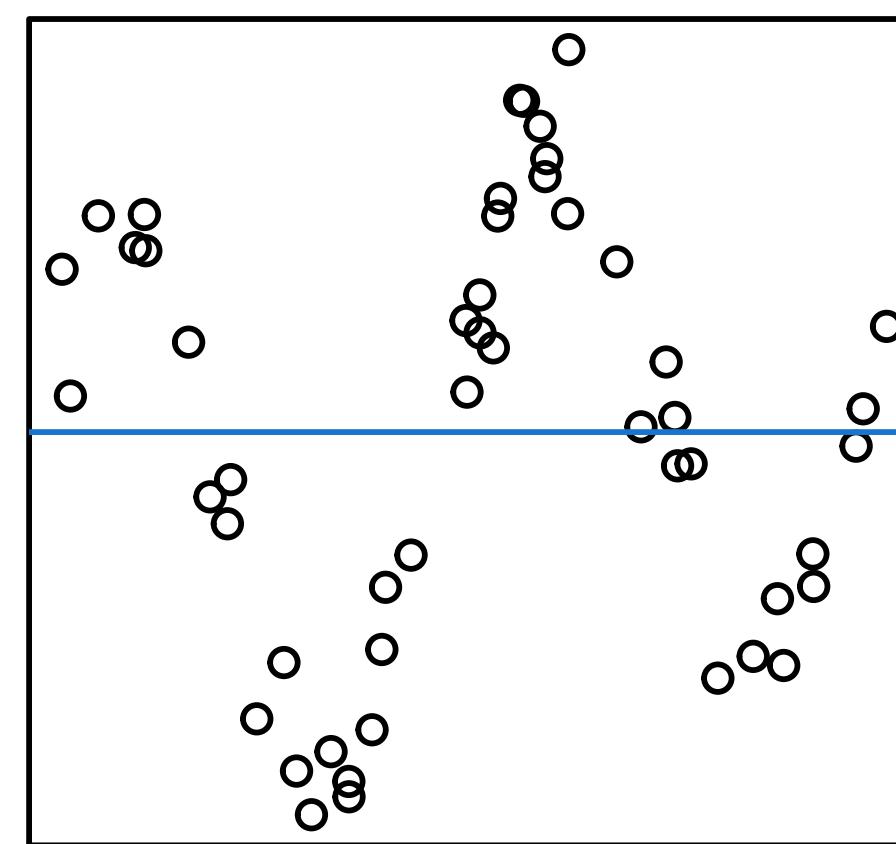
(11) A single positive outlier



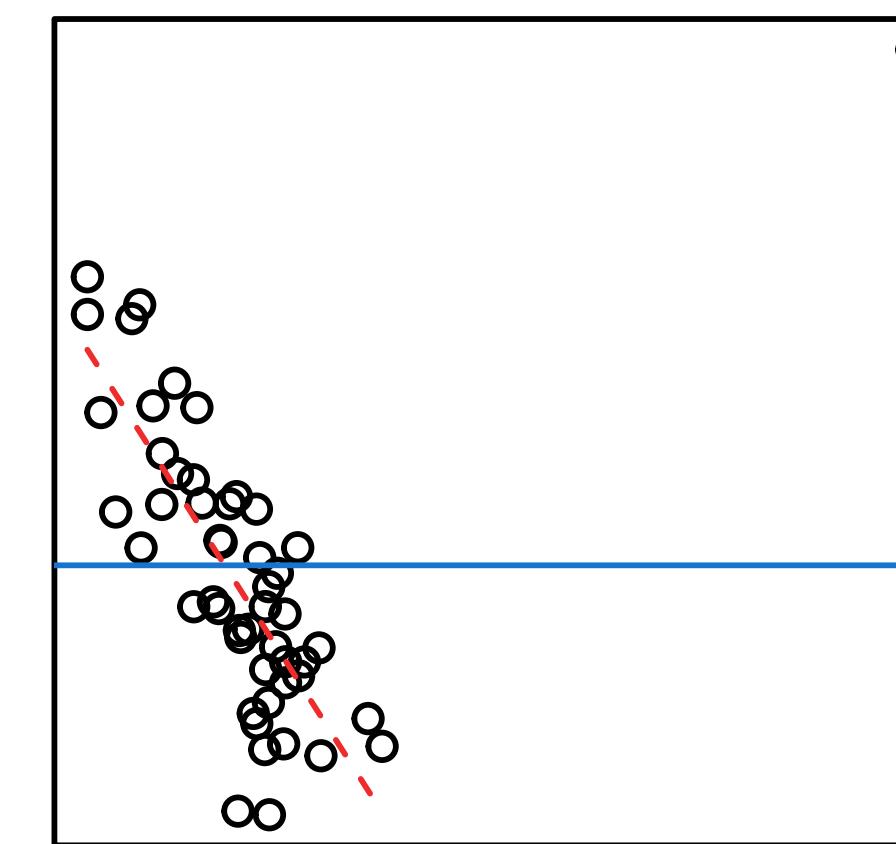
(12) A single negative outlier



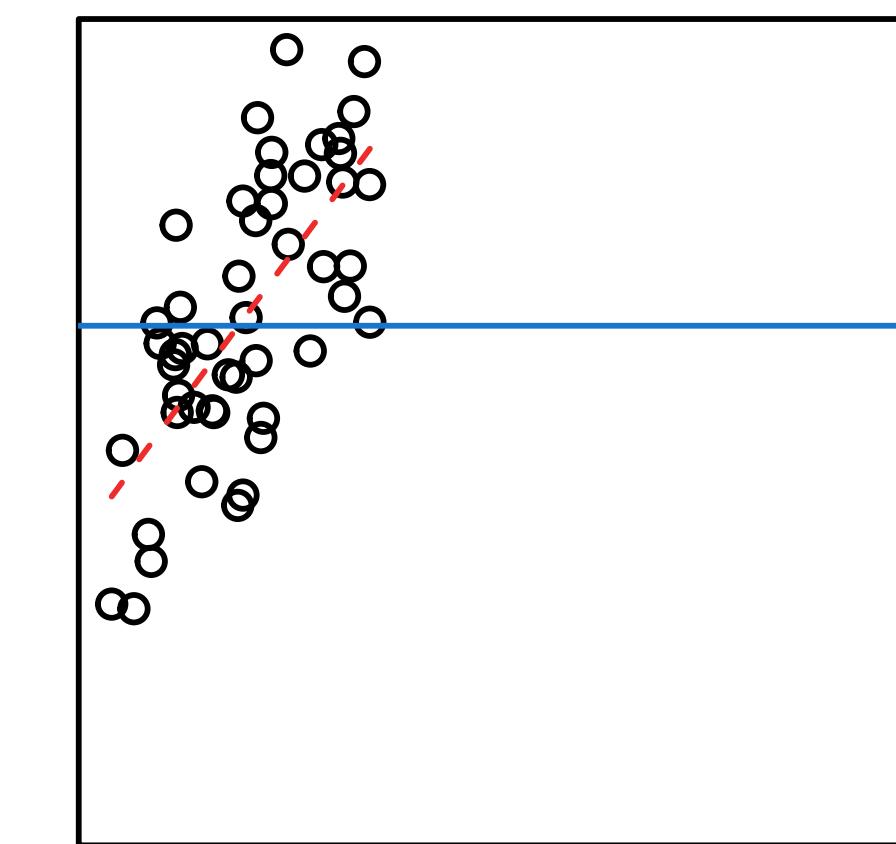
(13) Bimodal residuals



(14) Two groups

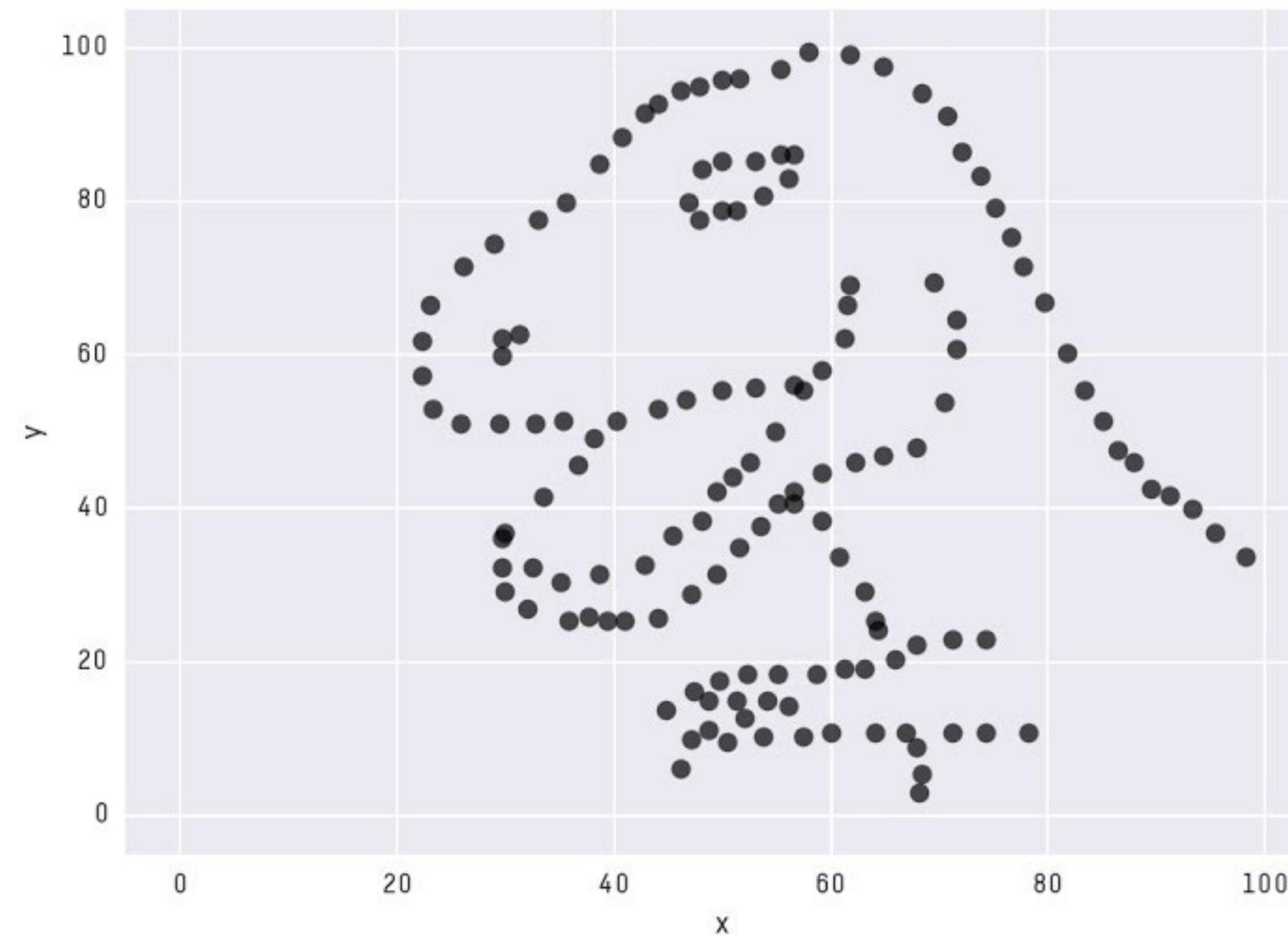


(15) Sampling at the extremes



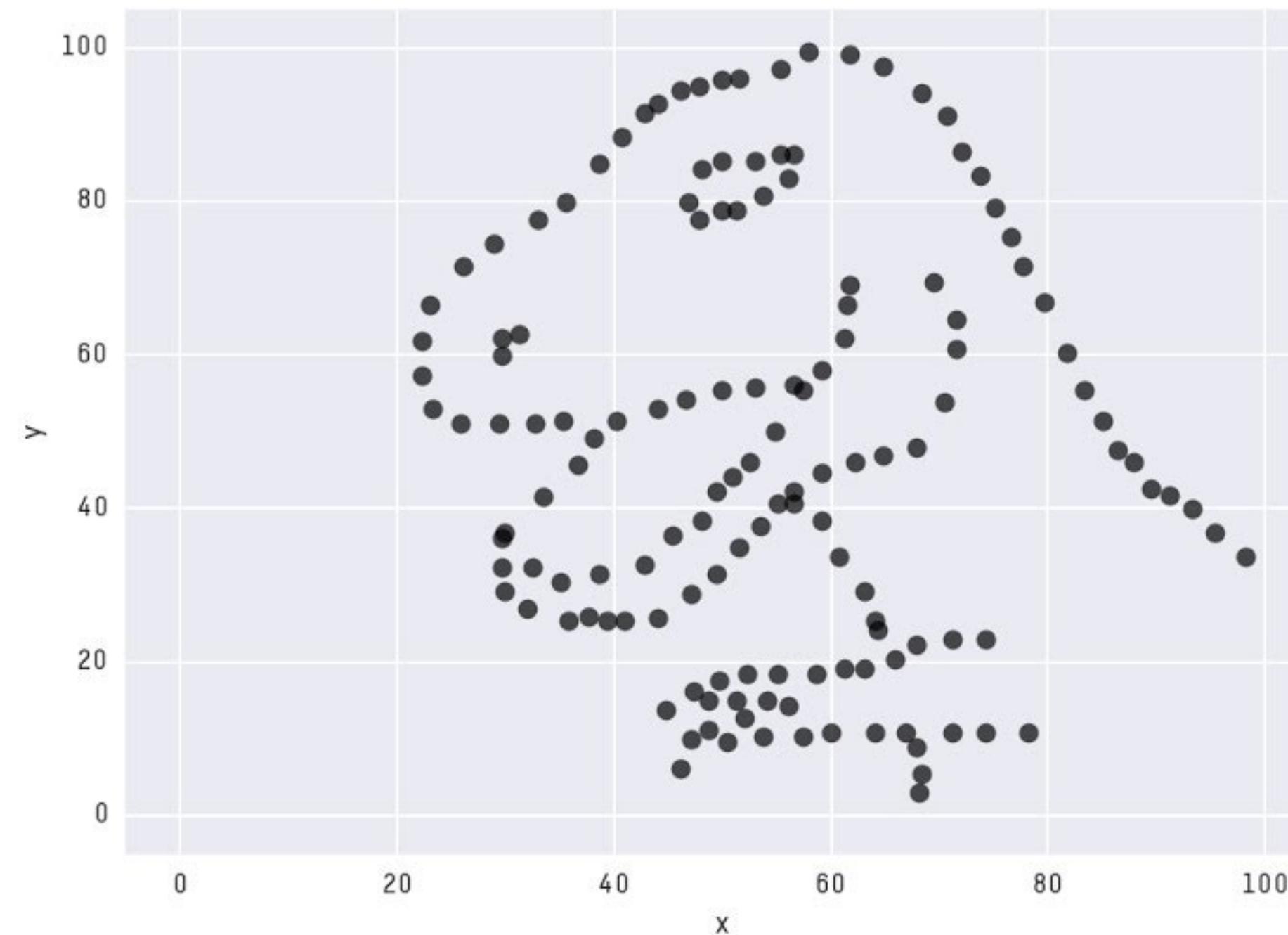
(16) Coarse data

<https://www.youtube.com/watch?v=DbJyPELmhJc>



X Mean: 54.2632025  
Y Mean: 47.8315781  
X SD : 16.7650109  
Y SD : 26.9353144  
Corr. : -0.0645195

<https://www.youtube.com/watch?v=DbJyPELmhJc>



X Mean: 54.2632025  
Y Mean: 47.8315781  
X SD : 16.7650109  
Y SD : 26.9353144  
Corr. : -0.0645195

# How many 5s?

192568719273163581623152957230  
519263912701749619236102701375  
069341629471037012639161

# How many 5s?

192 $\textcolor{red}{5}$ 68719273163 $\textcolor{red}{5}$ 816231 $\textcolor{red}{5}$ 29 $\textcolor{red}{5}$ 7230  
 $\textcolor{red}{5}$ 1926391270174961923610270137 $\textcolor{red}{5}$   
069341629471037012639161

**Our cognitive ability is limited.**

**Visual aids free up  
our mental capacity.**

Our visual system is a **massively parallel** pattern recognition machine.

# For the next class

- J. Heer et al. A Tour through the Visualization Zoo.  
<https://queue.acm.org/detail.cfm?id=1805128>
- J. VanderPlas, The Python Visualization Landscape. <https://youtu.be/FytuB8nFHPQ>
- Further readings: <https://github.com/yy/dviz-course/blob/master/docs/m01-intro/class.md>

# Questions?