

Meetup

THE 7TH SESSION
A SYSTEMATIC VIEW OF
DATA SCIENCE



BABAK REZAYI

ASSISTANT PROFESSOR AT DEPARTMENT
OF INDUSTRIAL ENGINEERING, FACULTY OF
ENGINEERING, FERDOWSI UNIVERSITY OF
MASHHAD

ICDS Innovation Center
for Data Science

اسلم داده

کاربرد تفکر سیستمی در

بابک رضایی

گروه مهندسی صنایع
دانشکده مهندسی دانشگاه فردوسی مشهد

۲۵ مرداد ۱۴۰۱

- تعریف سیستم یکپارچه سیستمی
- آشنایی با تفکر سیستمی
- معرفی برخی از موانع تفکر سیستمی
- موانع تفکر سیستمی در علم داده

- سیستم مجموعه ای از دو یا چند جزء است که سه شرط زیر را داشته باشد:
- رفتار هر جزء بر رفتار کل تاثیر دارد.
- رفتار اجزاء و تاثیر آنها بر کل، به هم وابسته اند.
- نحوه رفتار هر جزء و تاثیر آن بر کل، دست کم به نحوه رفتار يك جزء دیگر بستگی دارد. هیچ جزیی تاثیر مستقل بر کل سیستم ندارد.
- بدون توجه به نحوه تشکیل گروه های فرعی، هر کدام از گروه های فرعی، بر رفتار کل تاثیر دارند و تاثیر هیچ کدام از آنها مستقل نیست.
- اجزا سیستم چنان به هم متصلند که گروه های فرعی مستقل در آنها شکل نمی گیرد.

ویژگی های سیستم

- خاصیت‌های لازم هر سیستم به عنوان کل، از تاثیر متقابل اجزا آن ناشی می‌شود و نه از فعالیت‌های جداگانه اجزا.

- بنابراین وقتی سیستم به اجزا تفکیک می‌شود، خاصیت‌های ضروری خود را از دست می‌دهد.

سیستم کلی است که شناخت آن با تحلیل، امکان‌پذیر نمی‌باشد.

تفکر سیستمی

• دارای سه مرحله زیر است:

• تعیین يك مجموعه (سیستم) فراگیر به نحوی که آنچه که باید تشریح شود، بخشی از آن باشد.

• تشریح رفتار یا خواص مجموعه فراگیر

• تشریح رفتار یا خواص پدیده مورد نظر، بر حسب نقش‌ها و یا وظیفه‌های آن در مجموعه فراگیر

موانع تفکر سیستمی

- عدم اثربخشی الگوهای مکانیکی
- همه مشکلات سیستم را به بیرون از سیستم نسبت دادن
- تفکر بر اساس همبستگی بین عوامل به جای تفکر بر اساس رابطه علت و معلولی بین آنها
- من یعنی شغلم
- تمرکز بر روی وقایع
- پندار بیهوده در مورد یادگیری از تجارب

عدم اثر بخشی الگوهای مکانیکی

- کندي آسانسور

- برنامه ریزی خانواده در هند

همه مشکلات سیستم را به بیرون از سیستم نسبت دادن

- شرکت People Express Airlines

- تولیدکنندگان بنزین در آمریکا و تبلیغات افزایش فروش بنزین

«آن بیرون» و «این داخل» قسمت‌هایی از یک سیستم هستند.

همبستگی یا رابطه علت و معلولی

- ایکاف:
- یک متقال ادراک از رابطه علی، با ارزش تر از خروارها دانش درباره همبستگی است.
- میزان فروش بستنی و نرخ جنایتها
- همبستگی بین مصرف سرانه توتون در بیست و یک کشور با سرانه شیوع سرطان ریه
- مصرف سیگار در بیست و یک کشور فوق و همبستگی آنرا با سرانه شیوع وبا

من یعنی سلغم

- زمانی که از افراد سوال می‌شود که چه می‌کنند؟
- اغلب فعالیت‌های روزمره را برمی‌شمارند، نه اهداف اصلی مجموعه
- اکثر مردم خود را جزئی از سیستم می‌دانند که آنها یا بر روی آن اثری ندارند و یا تاثیرشان بسیار اندک است.
- در نتیجه آنها تمایل دارند که مسولیت‌هایشان را به محدوده شغلی خود محدود کنند.
- کارخانه‌های اتومبیل سازی آمریکا و ژاپن در طراحی شفت

تمرکز بر روی وقایع

- اتفاقات، وجه غالب صحبت‌ها در سازمان‌ها هستند:
- فروش ماه گذشته، کاهش جدید بودجه، درآمد فصل گذشته، چه کسی ارتقا یافته، محصول جدید شرکت رقیب و...
- تمرکز بر روی وقایع منجر به توضیح واقعه می‌شود، برخی از این تفاسیر به خودی خود نیز نادرست نیستند. اما، ما را از دریافتن الگویی تغییرات دراز مدت محروم می‌سازد.
- اصلی‌ترین تهدیدها که امروزه متوجه بقا سازمان‌ها هستند، نتیجه فرآیندهای آرام و تدریجی هستند و نه وقایع ناگهانی.

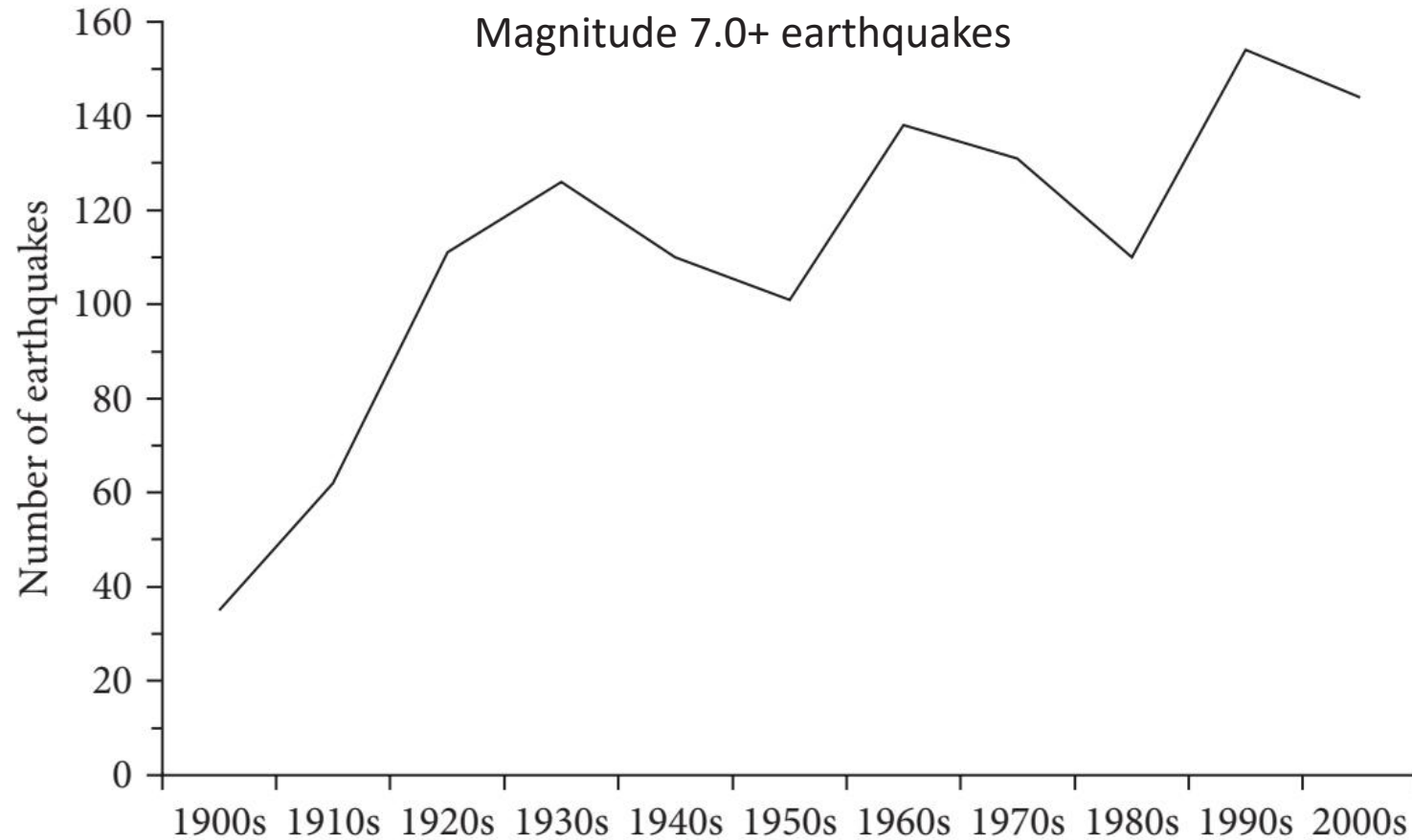
مركز بر رومي وقائع : فورمانه پخته

- صنایع خودروسازی آمریکا
 - در دهه 60 تسلط کامل
 - 1962 سهم ژاپنی‌ها کمتر از چهار درصد
 - 1967 سهم ژاپنی‌ها حدود ده درصد
 - 1974 سهم ژاپنی‌ها حدود پانزده درصد
 - اوایل دهه 80 سهم ژاپنی‌ها حدود 21.3 درصد
 - 1989 سهم ژاپنی‌ها حدود سی درصد

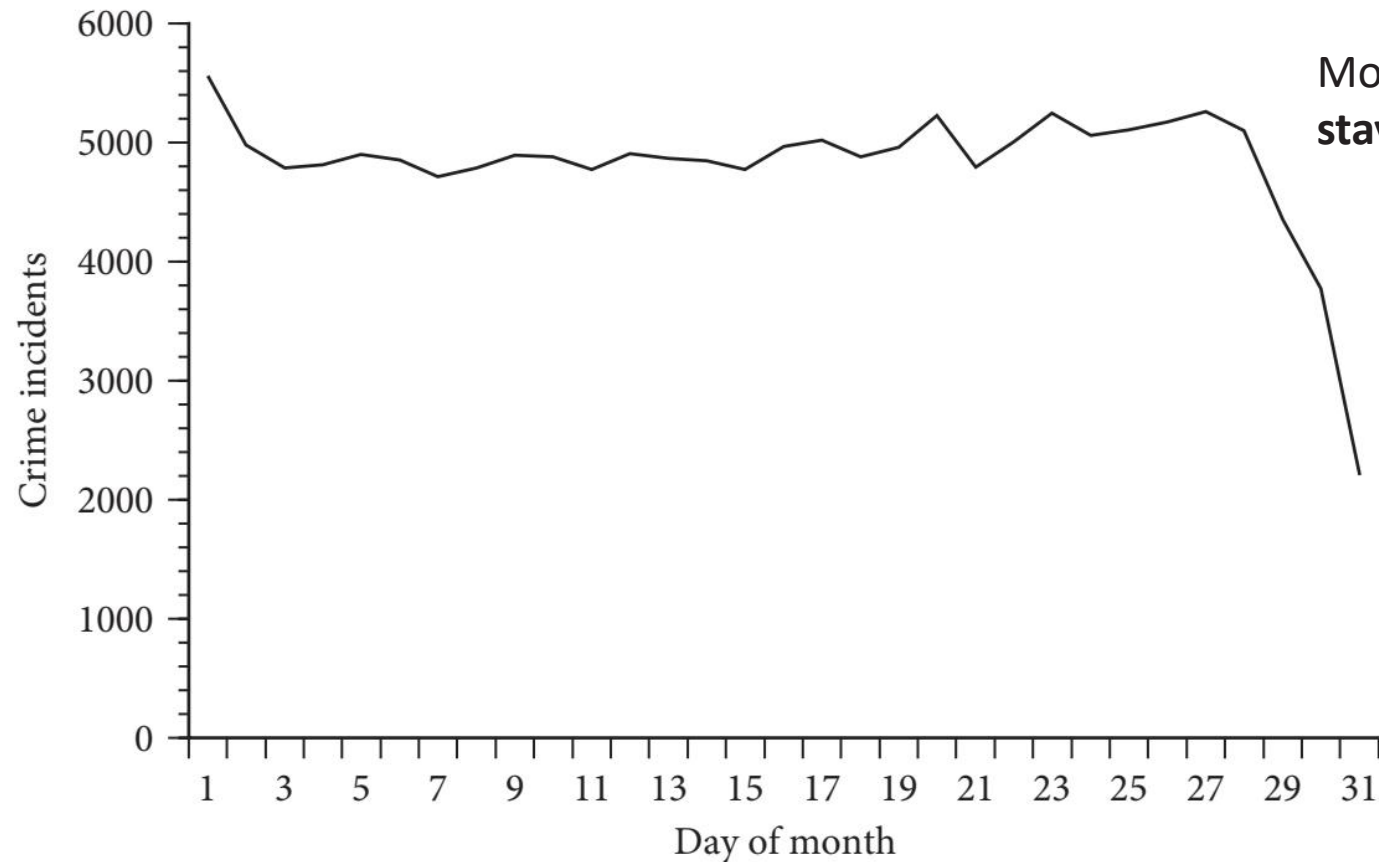
نذار پیسوده در مورد یادگیری از تجارب

- قوی ترین نوع یادگیری: تجربه مستقیم
- از طریق سعی و خطا و انجام عمل و مشاهده نتایج آن و تکرار عمل با توجه به نتیجه های قبلی
- اگر نتوانیم نتایج اعمال خود را درست تشخیص دهیم؟ یا نتایج آنها در آینده دور رخ دهد؟ چه باید کرد؟
- افق یادگیری:
- دامنه ای از زمان و مکان که در آن باید کارایی و ثمربخشی خود را مورد ارزیابی قرار دهیم.
- زمانی که نتایج اعمال ما دیرتر از «افق یادگیری» مشهود می شوند، دیگر نمی توان از طریق تجربه مستقیم چیزی را فرا گرفت.

Using Bad Data

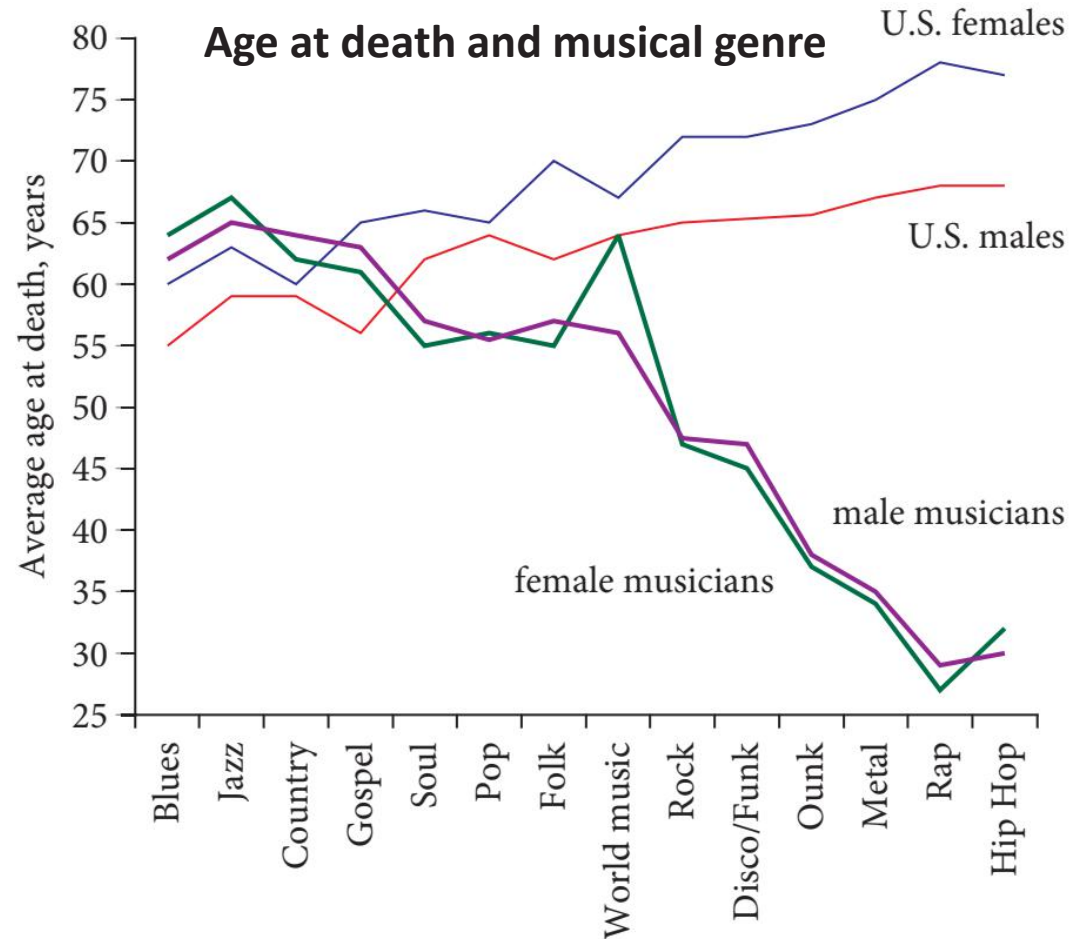


Using Bad Data



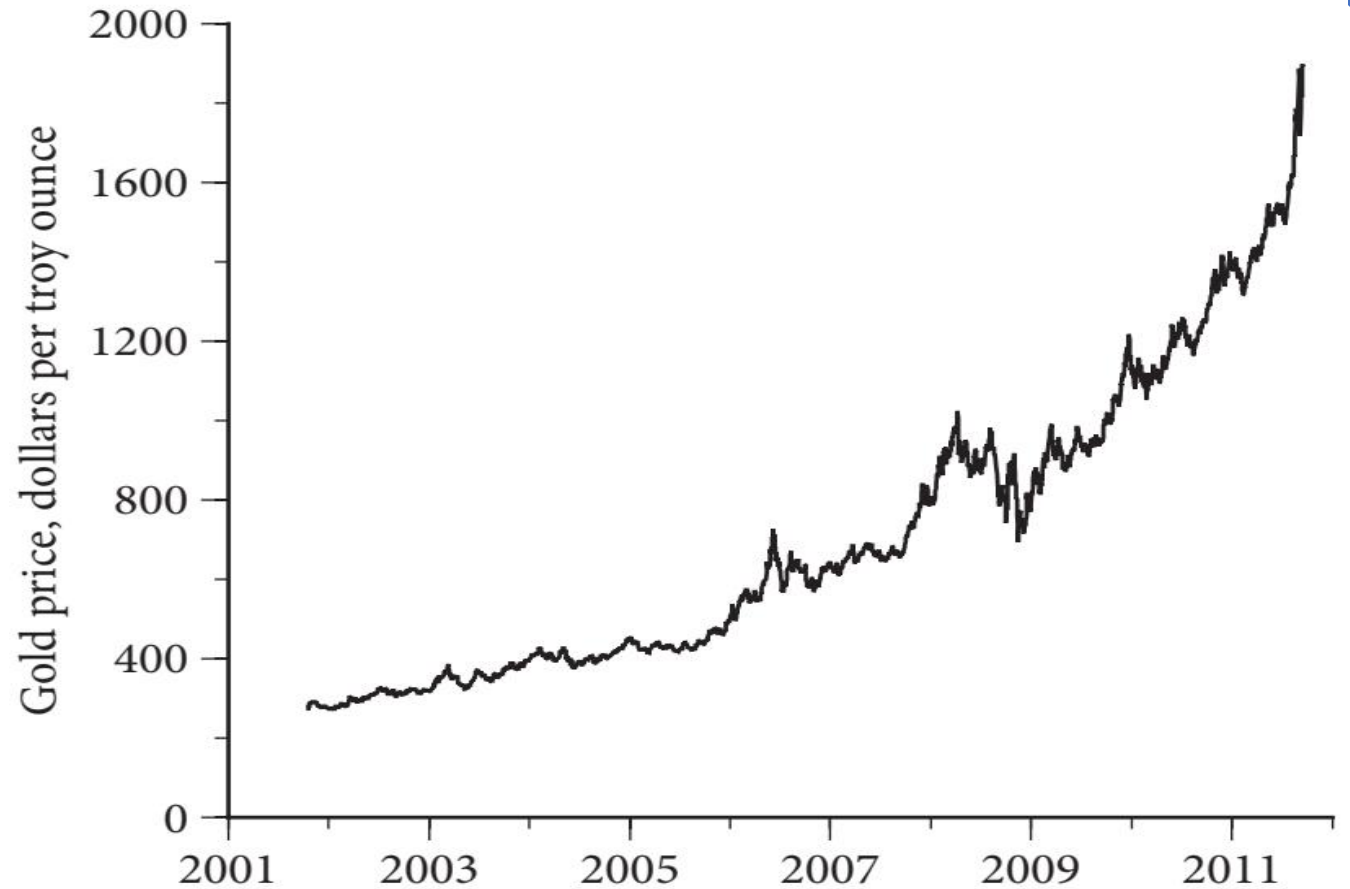
Most crime occurs on the first day of the month?
stay indoors on the first of the month

Using Bad Data



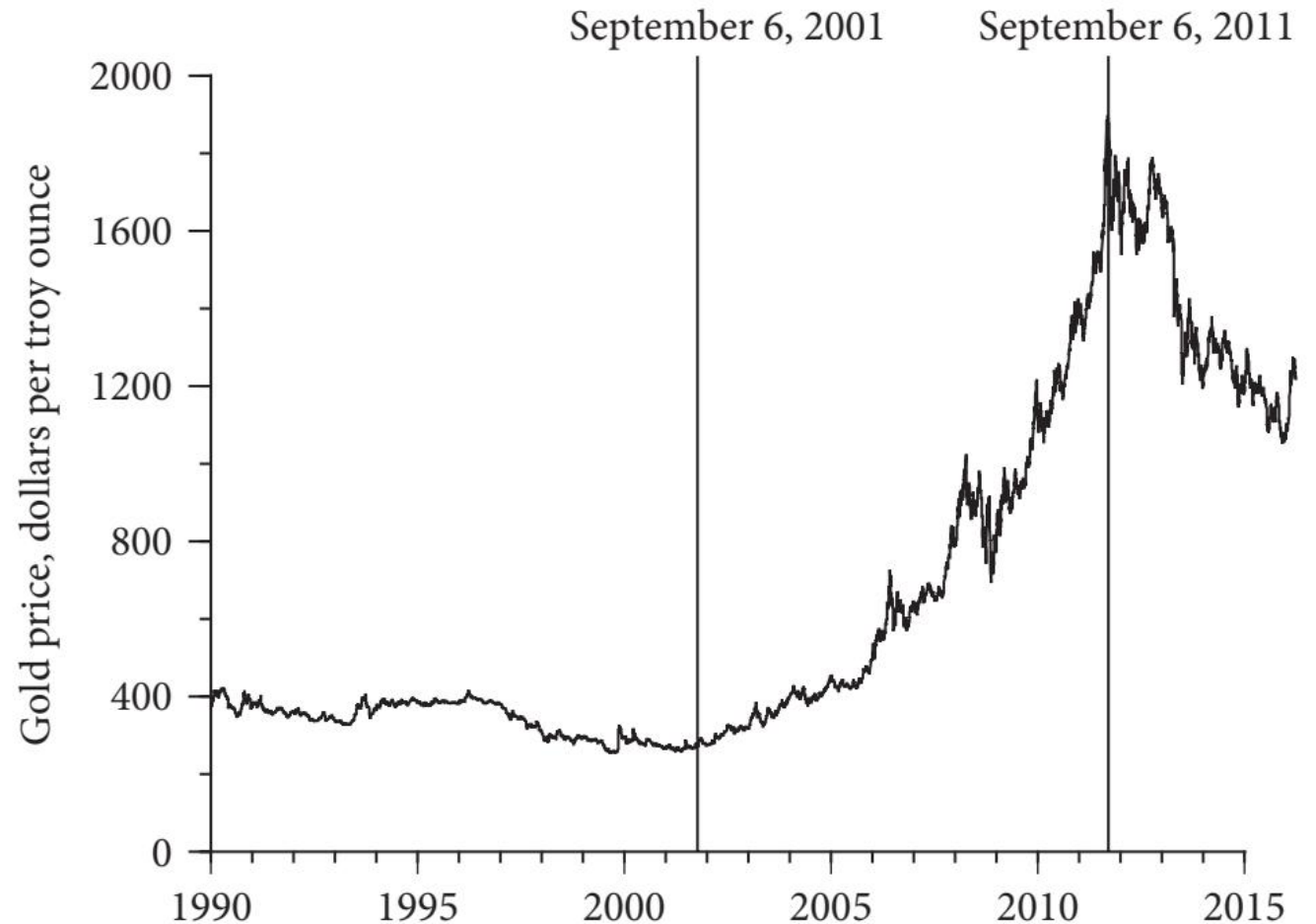
Using Bad Data

The April 4, 2016, issue of *Time* magazine contained a full-page advertisement touting gold as a can't-miss investment.

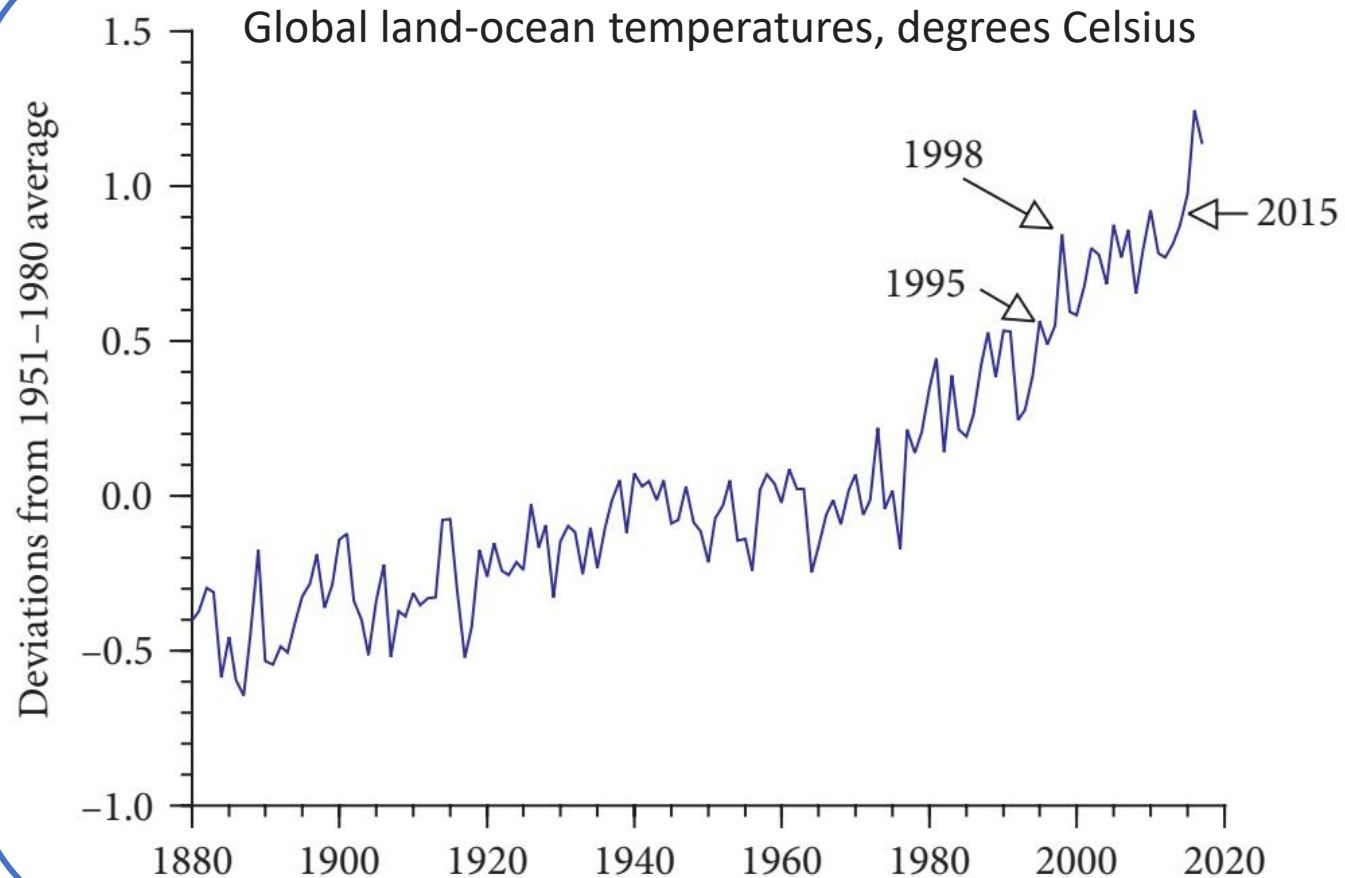


Using Bad Data

Good data scientists do not cherry pick data by excluding data that do not support their claims.



Using Bad Data



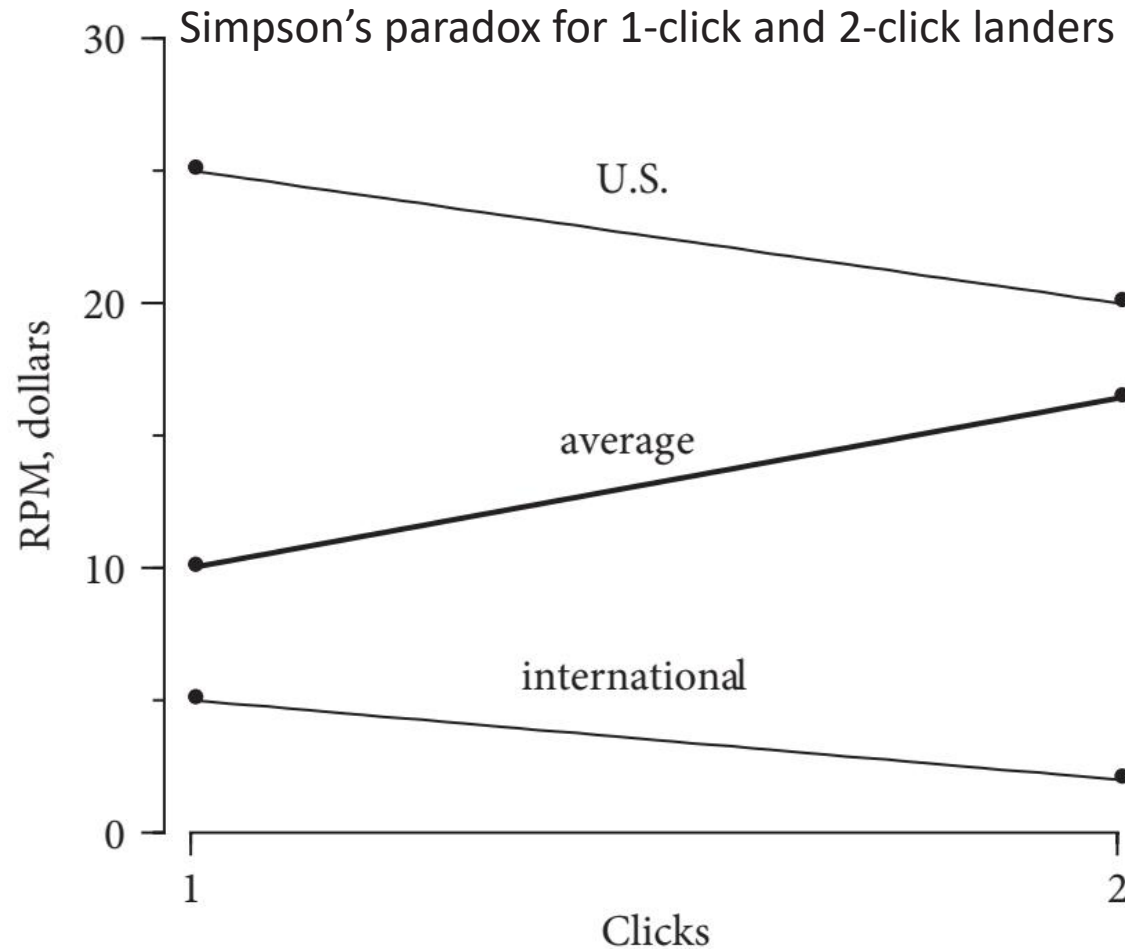
In a 2015 interview on “Late Night with Seth Meyers,” **Senator Ted Cruz** stated that, “Many of the alarmists on global warming, they’ve got a problem because the science doesn’t back them up. In particular, satellite data demonstrate for the last 17 years, there’s been zero warming.”

**IS YOUR
DATA
TRICKING
YOU ?**

SIMPSONS PARADOX



Simpson's paradox

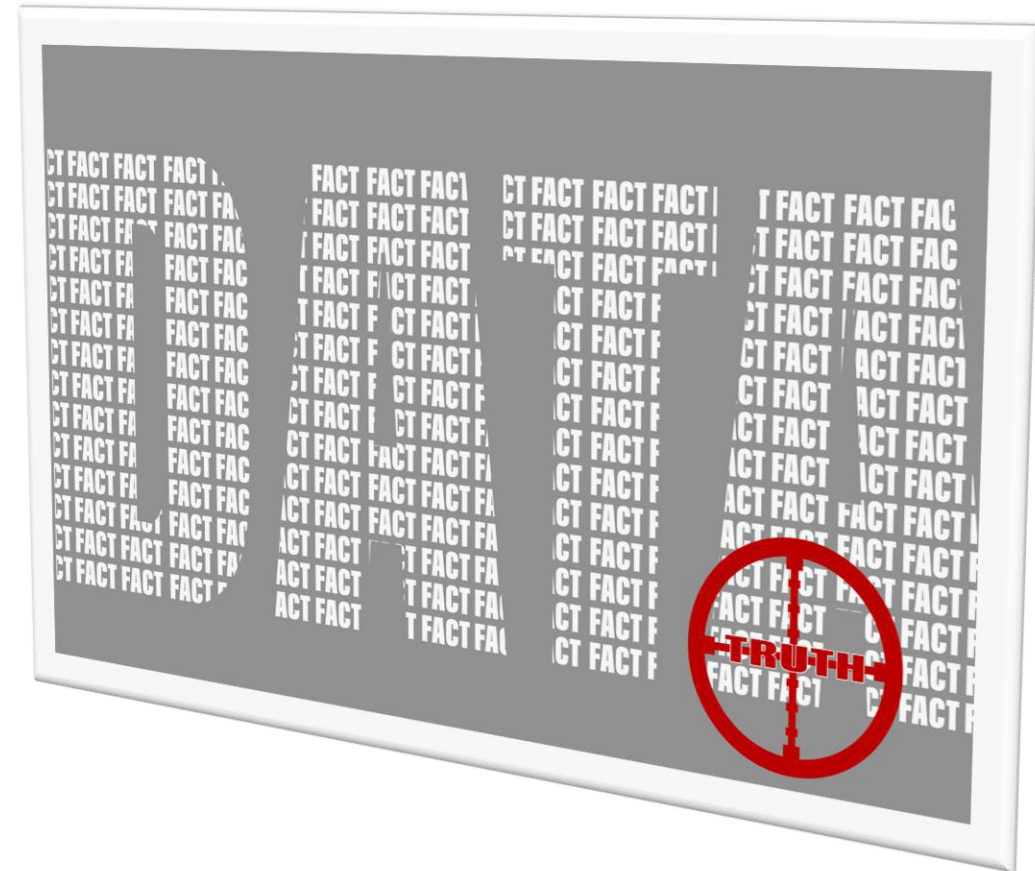


Simpson's Paradox occurs when aggregated and disaggregated data lead to contradictory conclusions.

Here, the aggregated data indicate that 2-click landers are superior, but when the data are disaggregated into United States and international, 1-click landers are superior.

Putting Data Before Theory

- Training Data & Test Data
- Variables



The Texas Sharpshooter Fallacy



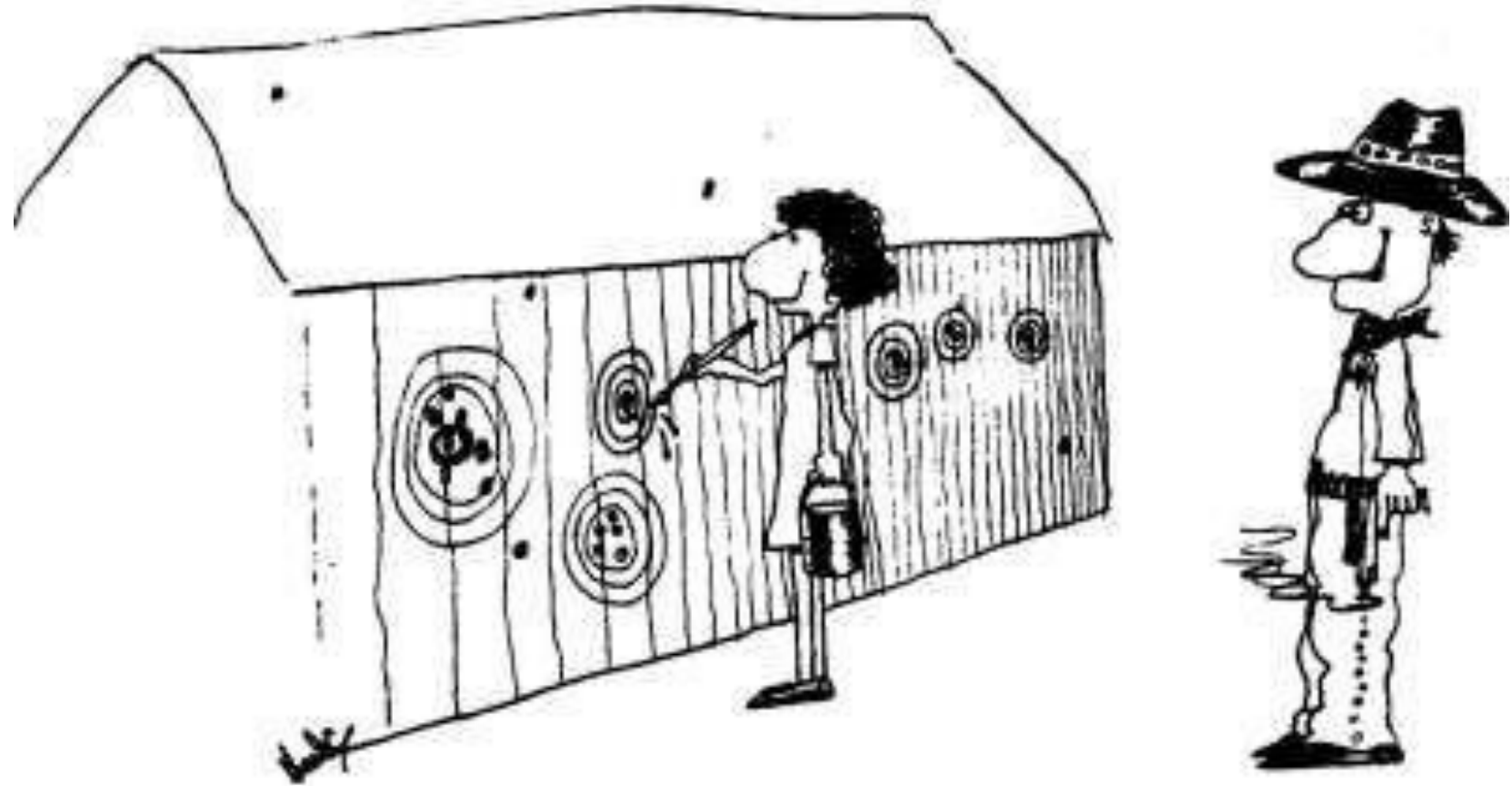
The Texas Sharpshooter Fallacy (1)

- A wanna-be cowboy paints hundreds of targets on a wall. When he fires his gun, he hits one of the targets, erases all evidence of the targets he missed, and proudly points to the target he hit as proof of his marksmanship. This evidence is worthless because, with so many targets, he was bound to hit one.
- In data science, this is analogous to doing hundreds (or thousands or millions) of tests and reporting only the ones with low p -values.

The Texas Sharpshooter Fallacy (2)

- The wanna-be cowboy shoots a bullet at a blank wall, and then draws a target around the bullet hole. This evidence of his marksmanship is worthless because the target was drawn after there was a bullet hole to draw the target around.
- In data science, this is equivalent to discovering a statistical pattern and concocting a theory to match the pattern.

The Texas Sharpshooter - Procedure



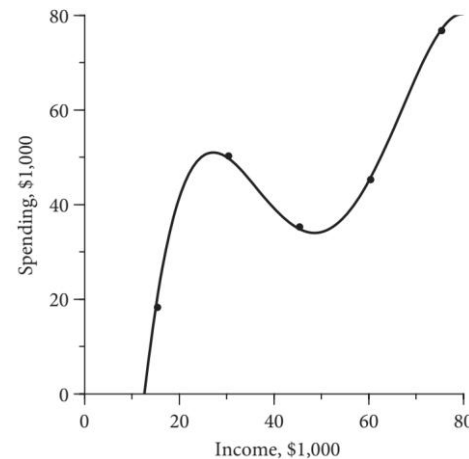
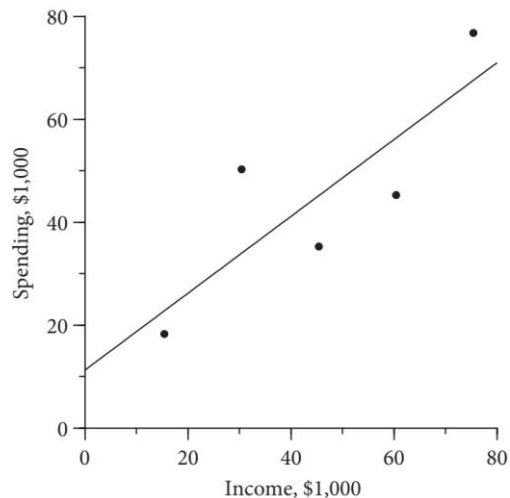
Putting Data Before Theory

We think that data patterns are unusual and therefore meaningful.
Patterns are, in fact, inevitable and therefore meaningless.

- This is why data mining is not usually knowledge discovery, but noise discovery.
- Finding correlations is easy. Good data scientists are not seduced by data-mined patterns because they don't put data before theory.

Worshiping Math

- “Garbage in, garbage out.”
- Good data scientists know that they need to get the assumptions right.
- Nonlinear models



Worshiping Math

- Good data scientists know that some predictions are inherently difficult and we should not expect anything close to 100 percent accuracy.
- It is better to construct a reasonable model and acknowledge its uncertainty than to expect the impossible.

- Instead of being blinded by math, good data scientists use assumptions and models that make sense.
- Good data scientists use math, but do not worship it.
- They know that math is an invaluable tool, but it is not a substitute for common sense, wisdom, or expertise.

Worshiping Computers

- There is a difference between using computers to do amazing things and trusting computers simply because they are capable of doing amazing things.
- Good data scientists use a variety of approaches to check and recheck their results.



کاربرد تفکر سیستمی در علم داده

بابک رضایی

گروه مهندسی صنایع
دانشکده مهندسی دانشگاه فردوسی مشهد

Torturing Data

Fooling Yourself

- Making unrealistic assumptions
- Measuring the wrong thing