

# 大数据时代的管理决策

*Lecture 0: Introduction to Big Data Analytics*

**Zhaopeng Qu**

**Business School, Nanjing University**

March.6 2022



# Outlines

- 1 Introduction: A Scientific Framework of Rational Knowledge
- 2 Big Data Creat Great Business
- 3 Who should take the course?
- 4 Logistics to the Course
- 5 Econometrics and Big Data Analytics
- 6 The Structure of Economic Data
- 7 Homework(课程研究项目)

# Introduction: A Scientific Framework of Rational Knowledge

# A Complex and Dynamic World



- Living in an unprecedentedly complex and dynamic world, we need **rational knowledge(理性认知)** and **scientific prediction(科学预测)**.

# Question #1: Student's Performance and Class Size



- A Classical Issue in Economics of Education: Student's Performance and Class Size
  - Is there a gap of students' performance between large-size classes and small-size classes?
  - Turn it into an empirical or policy question: What is the quantitative effect of reducing class size on student achievement?
  - Like by 5 student per class? or 10?

## Question #2: Discrimination in Employment



- Discrimination is the situation where people who are presumed to be equally productive or have equally productive capacity, get treated differently by the market.

## Question #2: Discrimination in Employment

- Because of their gender, their religion, sexual orientation or the color of their skin, whatever even though these characteristics are irrelevant for the purpose being considered.
  - **Racial Discrimination** in the market for housing loans.
  - **Gender Discrimination** in labor market
  - **Hukou Discrimination** in China
- To see if there is a difference in the probability of being denied **for otherwise identical** applicants and, if so, whether this difference is large or small.

## Question #3: Cigarette Taxes and Smoking



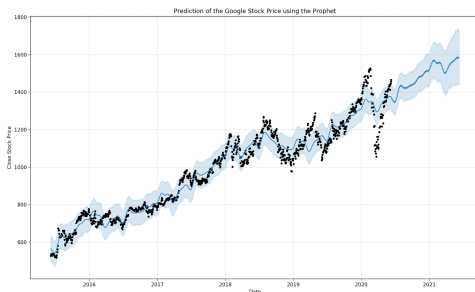
- Cigarette smoking is a major public health concern worldwide.
  - medical expenses of caring for those made sick by smoking.
  - nonsmokers breathe secondhand cigarette smoke.



## Question #3: Cigarette Taxes and Smoking

- Basic economics says: *cigarette prices go up, consumption will go down.*
  - But by how much? If the sales price goes up by 1%, by what percentage will the quantity of cigarettes sold decrease?
  - The percentage change in the quantity demanded resulting from a 1% increase in price is **the price elasticity of demand**, the numerical value, which economic theory can never provide.

# Question #4 How Will Stock Market Go?



- It seems that people always want a sneak preview of the future.
  - What will sales be next year at a firm that is considering investing in new equipment?
  - Will the stock market go up next month, and, if it does, by how much?
- A forecaster's job is to predict the future by using the past.
- Econometricians do this by using economic theory and statistical techniques to quantify relationships in historical data.

## Other Similar Questions

- Air pollution and Health?
- Credit regulation on housing price
- Coupon on products sales
- Trade War...
- Pandemic...

*Living in such a “messy” world, as a educated person, we  
rational knowledge and scientific prediction to make decisions.*

# How can we acknowledge rationally?

- Anecdotes(轶事) or Intuition(直觉)
- Theory(理论/逻辑推理)
  - Systematical methodology: Hypothesis, Logical deduction...
- Empirics (数据实证)
  - statistical inference from data.

# An Example: Smoke and Mortality

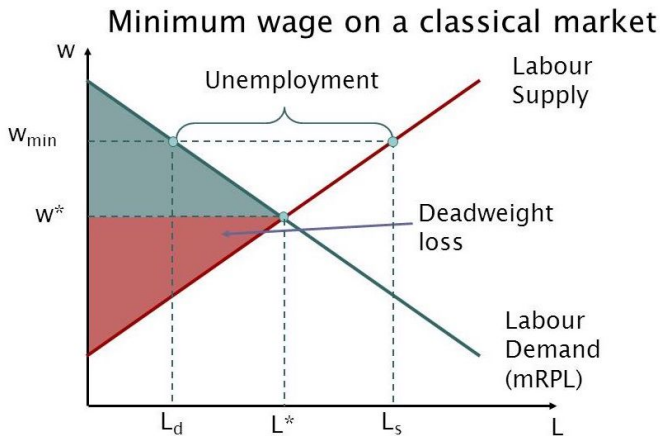
- Anecdotes(轶事) or Intuition(直觉)
  - eg. “My grandmother smoked two packs a day and lived until she was 95 years old.”
- Theory
  - 因为香烟中含有尼古丁、焦油等致癌物，所以...
- Empirics
  - 做实验或者通过抽样调查等方式收集数据，再用统计或计量方法来验证。

# Classical Theory Question: Human Capital v.s Signal

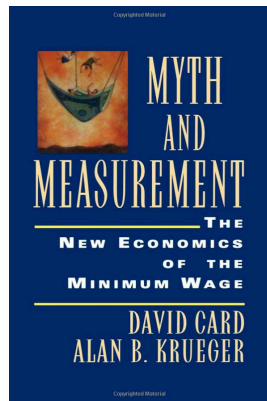
- A common phenomenon in labor markets can be observed all over the world.
  - **Higher education, Better pay!**
- Two classical theory to explain
  - **Human Capital:** Education improves work productivity.
  - **Signal:** Education does not increase the productivity. It simply serves as a signal of the individuals' innate ability.
- **Question: which one is right?**

# Hot Public Policy Debate: Minimum Wage and Unemployment

- Classical Supply-Demand Model tell us



# Hot Public Policy Debate: Minimum Wage and Unemployment

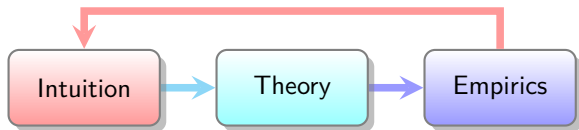


- One famous empirical evidence challenged the theory by Card and Krueger(1994)
- They found that increases in the minimum wage do **NOT** lead to job losses.



# An Scientific Workflow to Analyzing

- By Intuition: Propose meaningful or interesting questions(It does matter or we care about)
- By Theory: Obtain a preliminary conclusion or proposal an hypothesis
- By Empirics: use data and quantitative methods to test your theory or conclusion.



- *It is the job of **Econometrics or Causal Inference***

# Big Data Creat Great Business

Who should take the course?

# Who pursue an academic career

- Congratulations! You are in one of the most promising and internationalizing areas in economic research of China.
  - Master econometrics knowledge will improve your research greatly.
  - Your research is judged on how convincing it is.
  - Econometrics helps ensure and formalize credibility.
  - Overwhelming majority of top journal articles are quantitative.

# Who enter industry job market

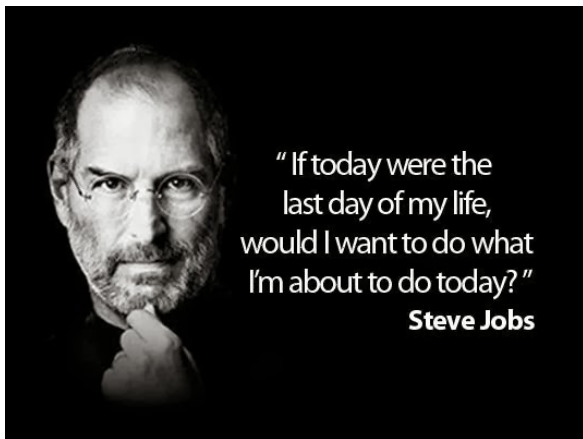
- Who want to work in industry: mastering it can help you get a good job!
  - A curious phenomenon: Why did the Wall Street hire more mathematics/statistics /economics graduates than anytime in history? Overwhelming financial data to analyze
  - Credit cards, POS terminals and ERP systems widely used in supermarkets, banks and factories also revolutionize the business (or management) practices in marketing, accounting, management operation etc.
  - A lot of internet giants even hire economists to lead their special R&D department. Such as
    - Google/Microsoft/eBay
  - Besides in finance, **Big data** are also grow vigorously in Consulting and Business areas.

# Who look for fun

- Introduction to Econometrics could not be a boring and demanding variant of a mathematics course, but **interesting and having fun**.
- You can just enjoy it by thinking in an empiricist' s way in your daily life.
  - have novel ideas or new perspectives about our world.
  - Econometrics is kind of a bible or philosophy of economists.
- We will cover many very interesting stories
  - Interesting Examples
    - Eg. **Crime and Abortion** in *Freakonomics* written by Steven Levitt.
    - Eg. What is **the value to be the president's son(or daughter)?** in *Economic Gangster* written by Raymond Fisman and Edward Miguel.

# Whoever and Whatever

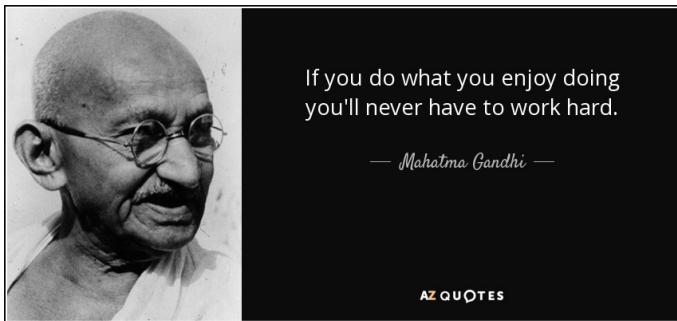
- **Whoever you would like to be or whatever you want**



- *Every choice you make has an opportunity cost, try your best to make a wise one.*

# Whoever and Whatever

- **Whoever you would like to be or whatever you want**



- *Enjoy doing something seriously and cultivate a special quality for yourself!*



# Hard and Soft Skills

- You **SHOULD** learn or improve several important skills during college.
  - **Hard Skills**
    - Language
    - Computer
  - **Soft Skills**
    - Critical Thinking
    - Presentation and Writing
    - Teamwork
- **Fortunately, you will learn/practice almost all above skills in our class.**

# Conclusion

- In a word, ***Big Data Analytics*** is a **very very very important** and **interesting** course,
  - Please think over before your take it!
  - Once you take it, please work hard on it!
  - And please enjoy working hard on it!

# Logistics to the Course

# About Me and our TA

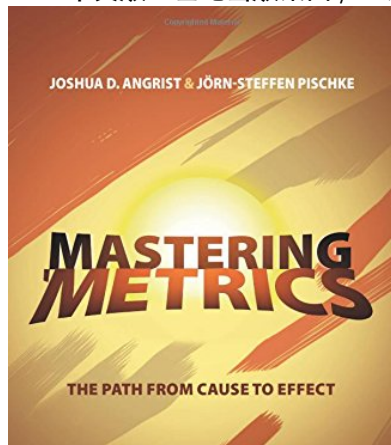
- My name is *Zhaopeng Qu*
  - Position and Affiliation: *Associate Professor, Institute of Population Studies, Business School.*
  - Research Fields: *Labor Economics and Applied Econometrics*
  - Office: Room 2017, Anzhong Building
  - Office Hour: Make an appointment in advance
  - Tel: 83621349
  - Email: [qu@nju.edu.cn](mailto:qu@nju.edu.cn)
- TA: Hongtao Wang
  - [graduate students](#)
  - good at *Stata*
  - Email:
- Online resources
  - Our Course Website: [\{\}https://byelenin.github.io/MBA\\_Big\\_Data/](https://byelenin.github.io/MBA_Big_Data/)
  - Dingding group: discuss anything about the course.

# Prerequisite

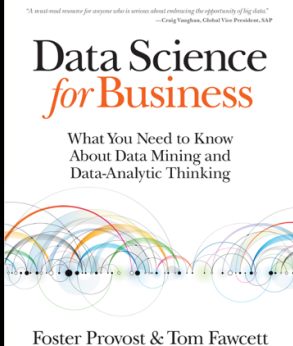
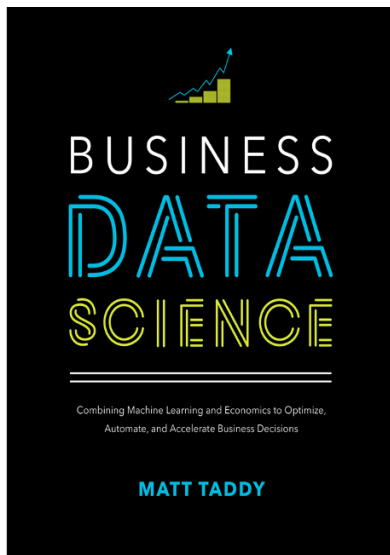
- Familiar with basic math(Probability and Statistics) (简单的数学基础)
- A minimal amount of computer programming experience. (接触过统计软件和简单编程)
- “Learning by Doing” is the best way to master knowledges and skills.

# Reference Textbooks

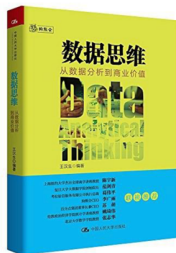
- Required Textbook: James H. Stock & Mark W. Watson, (2012). *Introduction to Econometrics*, 3rd Edition, Pearson Education.
  - 影印版: 格致出版社 / 上海人民出版社。
  - 中文版: 世纪出版集团 / 上海人民出版社。



## Business Data Science Textbooks



# Chinese Textbooks



- 王汉生,《数据思维：从数据分析到商业价值》，中国人民大学出版社，2017年9月。



# Computing Tools

- The main computing tools used in the course are ***Stata*** and ***R***, optionally.
  - ***Stata*** is a complete, integrated statistical package that provides everything you need for data analysis, data management, and graphics.
  - ***R*** is a free software and useful for statistical study and econometrics.

# Reference Books for Software

- Stata Instruction
  - 任何一本中文或英文书。
- R Instruction
  - 任何一本中文或英文书。
  - too many resources online such as *Datacamp*

# Interesting Books for Reading

- Steven D. Levitt and Stephen J. Dubner, *SuperFreakonomics: Global Cooling, Patriotic Prostitutes, and Why Suicide Bombers Should Buy Life Insurance*, 2009. (中译本, 《超爆魔鬼经济学》, 斯蒂夫·列维特、斯蒂芬·都伯纳著, 中信出版社, 2010年1月。)
- Ian Ayres, *Super Crunchers: Why Thinking-By-Numbers is the New Way To Be Smart*, 2007. (中译本《超级数字天才》, 伊恩·艾瑞斯著, 中国青年出版社, 2008年1月。)
- Viktor Mayer-Schonberger and Kenneth Cukier, *Big Data: A Revolution that will transform how we live, work and think*. (中译本《大数据时代: 生活、工作与思维的大变革》, 浙江人民出版社, 2013.1)
- Ian Ayres, *Super Crunchers: Why Thinking-By-Numbers is the New Way To Be Smart*, 2007. (中译本《超级数字天才》, 伊恩·艾瑞斯著, 中国青年出版社, 2008.1。)
- 崔小屹和韩青, 用数据说话——大数据时代的管理实践, 北京大学出版社, 2013年11月。(中国案例)

# What I promise to offer you

- Prepare lectures as well as possible.
- One to one interaction on topics covered in the course, especially for your own topics.
- Help you start to using Stata or R to analyze some popular data sets in China.
- A good score?
  - It depends on you.

# What I expect to you

- Class participation with a little bit aggressive attitude.
  - More questions, more scores!
  - Interrupt me as often as necessary! (but I know most of you are not comfortable to this)
  - Got a dumb question? Please assume that you are the smartest person in the class, and you eventually will be!
- Finish your homeworks.
- Self-motived learn by doing.

# Two iron rules

- Don't ever cheat on assignments and exams.



# Two iron rules



- **Don't ever snitch your teachers to help political persecution.**

# Evaluation

- Class Participation (20%)
- Team Project: A research proposal(80%)
  - Oral Presentation(40%)
  - Written Proposal (40%)



# Welcome contact me and TAs

- **Any Questions?**

# Econometrics and Big Data Analytics

# Introduction: Econometrics

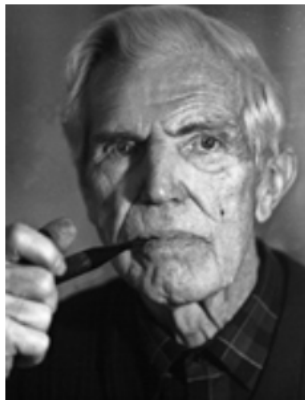


- The term is attributed to
  - **Ragnar Frisch(1895-1973);**
  - **1969 Nobel Prize co-winner**  
(the first year for Economics)
- Although the term coins by a combination of economics and metrology, it is special enough in social science and science at that time.

# Introduction: Econometrics

- “Econometrics is by no means the same as *economic statistics*. Nor is it identical with what we call general *economic theory*, although a considerable portion of this theory has a definitely quantitative character. Nor should econometrics be taken as synonymous with the application of *mathematics to economics*. Experience has shown that each of these three view-points, that of *statistics, economic theory, and mathematics*, is a necessary, but not by itself a sufficient, condition for a real understanding of the quantitative relations in modern economic life. It is the unification of all three that is powerful. *And it is this unification that constitutes econometrics.*”
  - (Ragnar Frisch, *Econometrica*, 1933, volume 1, pages. 1-2)

# Introduction: Econometrics



- ***Trygve Haavelmo(1911-1999)***
- **1989 Nobel Prize winner**
- “The method of econometric research aims, essentially, at a conjunction of **economic theory** and **actual measurements**, using the theory and technique of **statistical inference** as a bridge pier.”  
(*Econometrica*, 1944, volume 12, pages. 1-2)

# Introduction: Econometrics



- ***James Stock and Mark Watson(2014)***

- “Ask a half dozen econometricians what econometrics is—you could get a half dozen different answers.”
- “At a broad level, it is a **science and art** of using economic theory and statistical techniques to analyze economic data.”

# Introduction: Econometrics

- My View: **In general, a series of scientific methods to searching for economic logics from data.**
- It could include two jobs
  - *Making a causal inference*, such as
    - Testing economic theories.
    - Estimating causal effects.
    - Using data to give policy recommendations.
  - *Forecasting or predicting* future values
- More and more prevalence in
  - **other social science** such as **political science, sociology, law and education studies** etc
  - and **business practice**, like the hottest one: **Data Science.**

# Econometrics and methodological revolution in Social Science

- Social science (firstly started by Economics) is experiencing **two methodological “revolutions”** over the past few decades.
- On the one hand, there is the **“credibility revolution”**
  - A movement that emphasizes the goal of obtaining secure **causal inferences** (Angrist and Pischke, 2010)
- On the other hand, there is the **“Big Data revolution”**
  - A movement that emphasizes that how our increasing ability to produce, collect, store and analyze vast amounts of data is going to transform our understanding of the human affairs. (Schonberger, 2013)
- Obviously, Econometrics are playing a **key role** in these two revolutions.



# Big Data, Data Science and Econometrics

- The name comes from computer scientists working to do aggregation on data that is too big to fit on a single machine.
- **Data Science** is the umbrella term for **inference** in a world that is messier than in old statistic textbooks.
- **Big Data** is just *Applied Data Science*, especially focused on business and industrial applications.
  - Inference the structural models: Variable Selection and Dimension Reduction
  - prediction: Probabilistic Prediction and Classification Rules
- Big Data tools
  - partially straight out of previous statistics and econometrics classes (Cluster, Regression and Matching)
  - some are totally new (such as decision trees and neural networks).

# Big Data, Data Science and Econometrics

- There are many labels for what we do...
  - Econometrics
  - Statistics
  - Data Mining/Big Data/Data Science
  - Machine Learning(ML)
  - Artificial Intelligence(AI)
- Along this spectrum, you move from heavy focus on what things you are measuring (what real phenomena they correspond to) to a more practical 'useful is true' pattern discovery approach.
- *The similarities are much bigger than any distinctions.*

# The Structure of Economic Data

# Introduction

- Data Structure
  - Cross-sectional data
  - Time series data
  - Pooled cross-sectional data
  - Panel data

# 1. Cross-Sectional Data: (Main Focus)

- Units: individuals, households, firms, cities, states, countries, etc.
- Data on *multiple* agents at a *single* point in time

$$\{x_i, y_i, \dots\}_{i=1}^N; N = \text{Sample Size}$$

- Usually obtained by random sampling from the underlying population. It means

$$\{x_i, y_i \perp x_j, y_j\}, i \neq j \in N$$

- Cross-sectional data are widely used in economics and other social sciences:
  - labor economics
  - public finance
  - industrial economics
  - urban economics

# 1. Cross-Sectional Data: (Main Focus)

**TABLE 1.1** Selected Observations on Test Scores and Other Variables for California School Districts in 1999

Observation (District) Number	District Average Test Score (fifth grade)	Student-Teacher Ratio	Expenditure per Pupil (\$)	Percentage of Students Learning English
1	690.8	17.89	\$6385	0.0%
2	661.2	21.52	5099	4.6
3	643.6	18.70	5502	30.0
4	647.7	17.36	7102	0.0
5	640.8	18.67	5236	13.9
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
418	645.0	21.89	4403	24.3
419	672.2	20.20	4776	3.0
420	655.8	19.04	5993	5.0

*Note:* The California test score data set is described in Appendix 4.1.

•  $x = STRatio$ ;  $y = TestScore$ ;  $N = 420$

## 2. Time Series Data:(Not Cover)

- Observations on a variable (or several variables) over time, thus data on a *single* agent at *multiple* points in time

$$\{x_t, y_t \dots\}_{t=1}^T; T = \text{Sample Size}$$

- Examples:
  - stock prices
  - money supply
  - consumer price index(CPI)
  - gross domestic product(GDP)
  - automobile sales
- Economic observations can rarely be assumed to be independent across time. So we have to account for the dependent nature of economic time series.
  - Data frequency: minutes, hourly, daily, weekly, monthly, quarterly, annually.

## 2. Time Series Data:(Not Cover)

**TABLE 1.2** Selected Observations on the Growth Rate of GDP and the Term Spread in the United States: Quarterly Data, 1960:Q1–2013:Q1

Observation Number	Date (year:quarter)	GDP Growth Rate (% at an annual rate)	Term Spread (% per year)
1	1960:Q1	8.8%	0.6%
2	1960:Q2	-1.5	1.3
3	1960:Q3	1.0	1.5
4	1960:Q4	-4.9	1.6
5	1961:Q1	2.7	1.4
.	.	.	.
.	.	.	.
.	.	.	.
211	2012:Q3	2.7	1.5
212	2012:Q4	0.1	1.6
213	2013:Q1	1.1	1.9

*Note:* The United States GDP and term spread data set is described in Appendix 14.1.



### 3.Pool(Repeat) Cross-Sectional Data(Pseudo-Panels)(Not Cover)

- Pooled cross sections can be generated by combining **two or more** years cross-sectional data.
- Cross-sectional data in each year is independent with other years.
  - It means that data does not track the respondent multiple times.
  - But the data come from a same population in different time.
- For it has both cross-sectional and time series features, so allows consideration of changes in key variables over time.
- Simple pooling may also be used when the number of observations of a single cross section is small.
- It is widely used in:
  - Before-after comparisons of a government policy
  - Cohort studies
  - Cross-sectional analyses

### 3.Pool Cross-Sectional Data(Not Cover)

**TABLE 1.4 Pooled Cross Sections: Two Years of Housing Prices**

obsno	year	hprice	proptax	sqrfc	bdrms	bthrms
1	1993	85500	42	1600	3	2.0
2	1993	67300	36	1440	3	2.5
3	1993	134000	38	2000	4	2.5
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
250	1993	243600	41	2600	4	3.0
251	1995	65000	16	1250	2	1.0
252	1995	182400	20	2200	4	2.0
253	1995	97500	15	1540	3	2.0
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
520	1995	57200	16	1100	2	1.5

## 4. Panel(or Longitudinal) Data

- Time series for each cross-sectional member in the data set, thus data on multiple agents at multiple points in time.
- The same cross-sectional units (individuals, firms, countries, etc.) are followed over a given time period.

$$\{X_{it}, Y_{it} \dots\}_{i=1, t=1}^{NT}$$

- Advantages of panel data:
  - Controlling for (time-invariant) unobserved characteristics
  - Consideration of the effects of lag variables

## 4.Panel(or Longitudinal) Data

**TABLE 1.3** Selected Observations on Cigarette Sales, Prices, and Taxes, by State and Year for U.S. States, 1985–1995

Observation Number	State	Year	Cigarette Sales (packs per capita)	Average Price per Pack (including taxes)	Total Taxes (cigarette excise tax + sales tax)
1	Alabama	1985	116.5	\$1.022	\$0.333
2	Arkansas	1985	128.5	1.015	0.370
3	Arizona	1985	104.5	1.086	0.362
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
47	West Virginia	1985	112.8	1.089	0.382
48	Wyoming	1985	129.4	0.935	0.240
49	Alabama	1986	117.2	1.080	0.334
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
96	Wyoming	1986	127.8	1.007	0.240
97	Alabama	1987	115.8	1.135	0.335
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
528	Wyoming	1995	112.2	1.585	0.360

# Corresponding sub-fields or sub-course

- Micro-Econometrics
  - Cross-Sectional
  - Pool Cross Sectional
  - Short Panel(large N, small T)
- Macro-Econometrics
  - Times series
  - Long Panel(small N, large T)

# Source

- Traditional Collecting Way:
  - Survey(调查)
  - Administrative data(官方业务数据)
- Collecting Data in Digital Times:
  - Mass or Big data:
    - Online documents
    - Social Media
    - Geolocations or Geographic data
  - High Frequency Data
    - Stock, future or other financial tractional data

# Data in China

## ● Survey Data

- China Household Income Project(CHIP)
- China Family Panel Survey(CFPS)
- China Health and Retirement Longitudinal Study(CHARLS)

## ● Administrative data:

- Census: 全国人口普查数据; 全国 1% 人口抽样调查;
- 工业企业数据库;
- 海关交易数据库;

## ● Online data:

- Taobao, JD, Tmall
- Movie Data on Douban.com(豆瓣电影数据)
- Air Quality: PM2.5(空气质量数据)
- Night-Lights Data(夜间灯光数据)
- Land Transaction Markets(土地交易市场数据)

# Big Data: Examples

## Big Data Application examples in different Industries:

### Retail/Consumer

- ❖ Merchandizing and market basket analysis
- ❖ Campaign management and customer loyalty programs
- ❖ Supply-chain management and analytics
- ❖ Event- and behavior-based targeting
- ❖ Market and consumer segmentations

### Finances & Frauds Services

- ❖ Compliance and regulatory reporting
- ❖ Risk analysis and management
- ❖ Fraud detection and security analytics
- ❖ Credit risk, scoring and analysis
- ❖ High speed arbitrage trading
- ❖ Trade surveillance
- ❖ Abnormal trading pattern analysis

### Web and Digital media

- ❖ Large-scale clickstream analytics
- ❖ Ad targeting, analysis, forecasting and optimization
- ❖ Abuse and click-fraud prevention
- ❖ Social graph analysis and profile segmentation
- ❖ Campaign management and loyalty programs

### Health & Life Sciences

- ❖ Clinical trials data analysis
- ❖ Disease pattern analysis
- ❖ Campaign and sales program optimization
- ❖ Patient care quality and program analysis
- ❖ Medical device and pharmacy supply-chain management
- ❖ Drug discovery and development analysis

### Telecommunications

- ❖ Revenue assurance and price optimization
- ❖ Customer churn prevention
- ❖ Campaign management and customer loyalty
- ❖ Call detail record (CDR) analysis
- ❖ Network performance and optimization
- ❖ Mobile user location analysis

### Ecommerce & customer service

- ❖ Cross-channel analytics
- ❖ Event analytics
- ❖ Recommendation engines using predictive analytics
- ❖ Right offer at the right time
- ❖ Next best offer or next best action



# Big Data: Volume

- Big in volume in terms of
  - the number of observations(size  $n$ )
  - the number of variables(dimension  $p$ )
- Eg. Wechat(微信) in 2018
  - 1.082 billion active users every month
  - 45 billion messages every day
  - 0.41 billion video calls every day

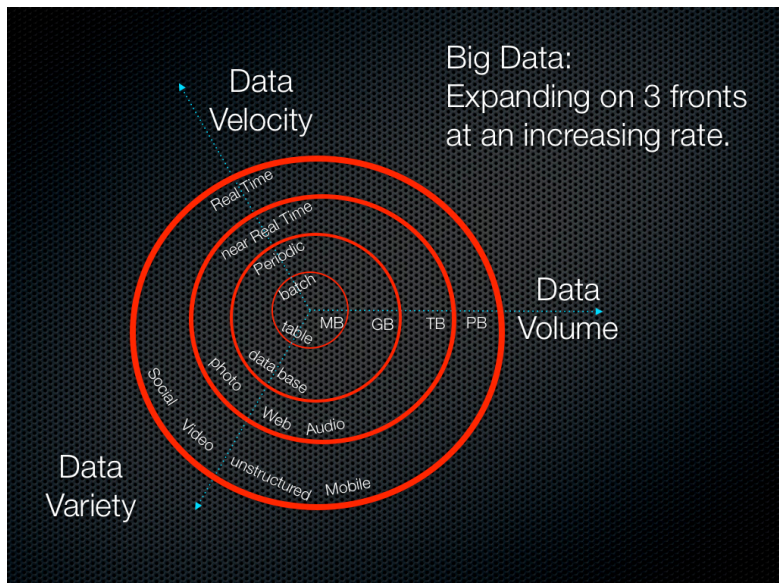
# Big Data: Variety

- Big in variety in terms of
  - classical data(numbers in tables)
  - photo/audio/video
  - text
  - map
  - sensors

# Big Data: Velocity

- Velocity is the measure of how fast the data is coming in.
  - Eg. Facebook(脸书) users upload more than 900 million photos a day in 2016
  - Eg. Wechat(微信) in 2018
    - 45 billion messages every day
    - 0.41 billion video calls every day

# Big Data: 3Vs



# Big Data means to analytics

- Big in the number of observations: size  $n$

$$n \rightarrow +\infty$$

- Big in the number of variables: dimension  $p$

$$p \rightarrow +\infty$$

- **Big Data** is just about inferring useful signal at massive scale.
  - Inference the structural models: Variable Selection and Dimension Reduction
  - Prediction: Probabilistic Prediction and Classification Rules.
- Tools in Big Data are
  - partially straight out of previous statistics and econometrics classes(Such as RCT, Regression and Matching)
  - some are totally new( such as trees and neural network).
- In the next lecture, I will introduce some basic concepts and methods in Big Data Analytics
  - Causal Inference(Econometrics)
  - Prediction(Machine Learning)

# Homework(课程研究项目)

# Homework(期末研究项目的准备工作)

- 课程结业项目分组：以自愿为原则，按照合理人数分组。
- 收集合适的的数据：最好能够使用自己公司相关的业务数据。
- 确定数据来源、收集方式、样本量等基本信息，判断该数据属于哪种数据结构。
- 下周日的课堂上，向我汇报研究项目的基本设计
  - 题目与内容
  - 变量（自变量、因变量、控制变量）
  - 方法（因果推断、预测）