Computational social science

An Introduction II

Hello World and Computational Social Science

Week 02

GSIAS, HUFS, Data Mining for Social Science

Qualitative and Quantitative issues

Quantitative

- Research knows the research target
- The research should be fully designed
- Availability of the data, e.g.) economic and political indicators, survey data or other equivalent numerical data.
- Is it possible to transform its data into numerical?
- Is it EMPIRICAL?
- Testing the Hypothesis.

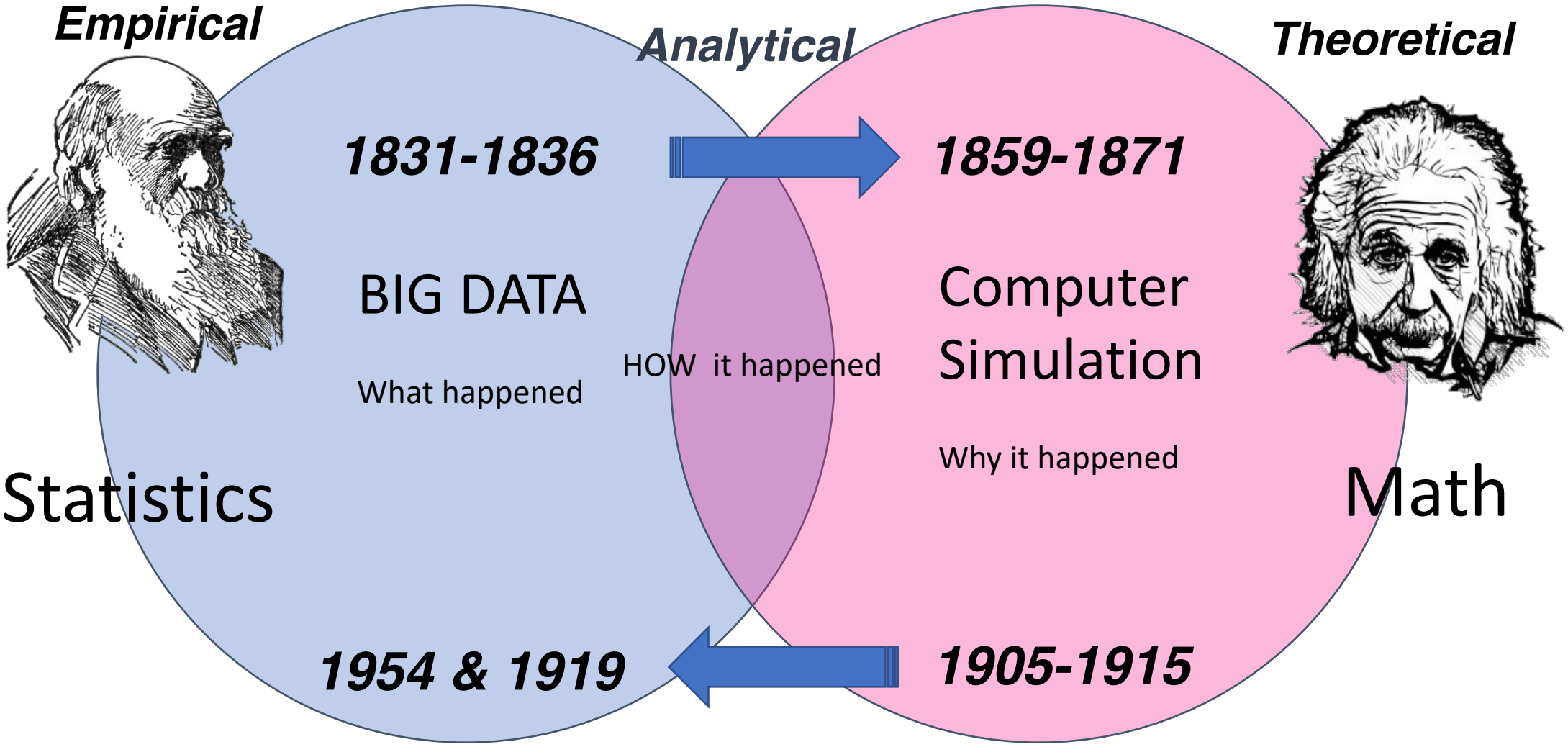
Qualitative

- Describing **"the big picture"**
- May only roughly know the research question and is in the preliminary stage.
- **However, if the researcher is confident to give a completely detailed description.**
- Data may be pictures, community, and literature, which can be interpreted with participation.
- It can be more rich information than Quantitative.

Observers' paradox

- I need to go into community A to conduct the research
- But by going into Community A, How can I be sure I am not influencing or altering the behaviour of community A by simply being present?
- But, If I do not go, I can not get the information.

The Scientific Method



The Scientific Method

USING PAST EXPERIENCE TO MAKE PREDICTION

Induction

Data to Theory: Data is Abundant, and good theory is scarce

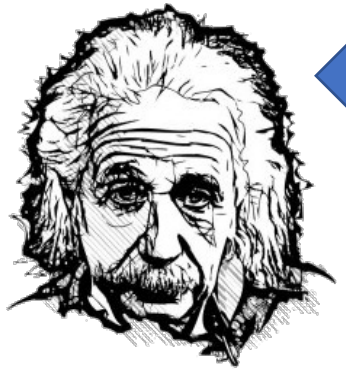
Phenomena

Data Collection
Observation
(Record, experiment, survey, simulation)

Abstraction
Analysis
Modelling

Data Collection
Observation
(Record, experiment, survey, simulation)

Theory



Deduction

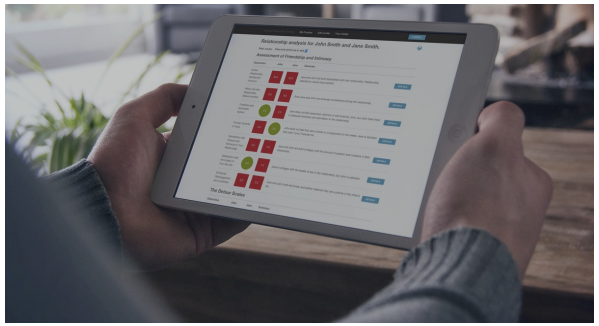
Theory to Data: Thoughts are Abundant, and good data is scarce

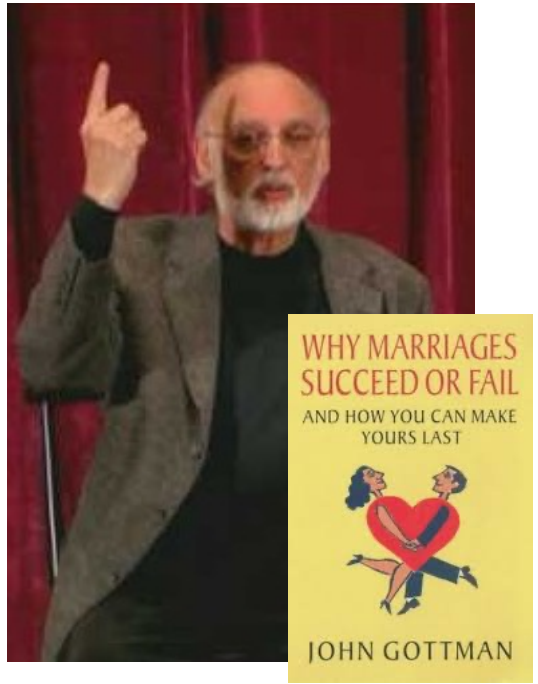
The Scientific Method: Limitation of Induction

The Scientific Method: Limitation of Deduction

Theory to Phenomena?

1974





John M. Gottman

Basic Grammar of R

The simple cookbook for R

Looking At Some of the Unique Features of R

```
# assign the values 1:5 to a vector called x:
x <- 1:5
x
[1] 1 2 3 4 5

# add the value 2 to each element in the vector x:
x + 2
[1] 3 4 5 6 7

# add one vector to another. To add the values 6:10 elementwise to x, you do the following:
x + 6:10
[1] 7 9 11 13 15
```

Exploring R

To start working in R, you need two things.

An Interface: You also need an interface to send that code to R.

https://cran.r-project.org
Download R for Linux (Debian, Fedora/Redhat, Ubuntu)
Download R for macOS
Download R for Windows

An Editor: you need an easy tool to write and edit code.

https://www.rstudio.com
https://www.rstudio.com/products/rstudio/download/

Saying ”Hello” to the world

```
# Hello Word
print("Hello World")
[1] "Hello World"

# Doing simple math
# Addition
 7+3
[1]  10

# Subtraction
7-3
[1]  4

# Multiplication
8*7
[1]  56
```

```
#Division
100/50
[1]  2

# Square root
sqrt(81)
[1]  9

# Subtraction
7-3
[1]  4

# Exponents
9^2
[1]  81
```

Using Vectors

A vector is the simplest type of data structure in R.
The R manual defines a vector as "a single entity consisting of a collection of things"

```
# To construct a vector, type as follow:
# c(1,2,3,4,5) 'c' stands for combine.
c(1,2,3,4,5)
[1] 1 2 3 4 5

# Operate sequential vectors
1:5
[1] 1 2 3 4 5

# Assigning vector
# to assign a vector with the '<-' as the operator.

x <- c(1,2,3,4,5)
x
[1] 1 2 3 4 5
```

```
# it works on numeric and text data
h <- "Hello"
h
[1]   Hello

# Furthermore, it is able to combine texts
hw <- c("Hello", "world!")
Hw
[1] "Hello" "world!"

# If needs to be a concatenation with words
paste("Hello", "world!")
[1] "Hello world!"
```

Vectorizing your functions

The lecture recommends using a google style R guide

https://web.stanford.edu/class/cs109l/unrestricted/resources/google-style.html

```
# Vectorized functions
Test.scores <- c(80, 85, 100, 90, 75)
Test.scores
[1]  80  85 100  90  75

# sum, mean, median, and mode
sum(Test.scores)
[1] 430
mean(Test.scores)
[1] 86
median(Test.scores)
[1] 85
mode(Test.scores)
[1] "numeric"
```

Miscellaneous information

```
# List the environments
ls()
[1] ”x”, ”h”, ”hw”

# remove an environment
rm(x)
ls()
[1] ”h”, ”hw”
```

```
# Installing the Package
install.packages("quanteda")

# load the Package
library(quanteda)
```

```
# Import Dataset
# Real Excel File
library(readxl)
Sample01 <-read_excel(file.choose())
View(Sample01)

# Real CSV File
Sample01 <-read.csv(file.choose())
View(Sample01)
```

Manipulating data frame

Creating data frame

```
# Creating Data Frame
ls()
[1] ”x”, ”h”, ”hw”

# remove an environment
rm(x)
ls()
[1] ”h”, ”hw”
```

```
# Installing the Package
install.packages("quanteda")

# load the Package
library(quanteda)
```

```
# Import Dataset
# Real Excel File
library(readxl)
Sample01 <-read_excel(file.choose())
View(Sample01)

# Real CSV File
Sample01 <-read.csv(file.choose())
View(Sample01)
```

Homework (Recommend)

Please Practice for your own benefit.

```
# Princeton Univ R Practice Library
install.packages("swirl") # install the package
library(swirl) # load the package

# Take a practice
library(swirl
swirl()

# More information about swirl is available at http://swirlstats.com/
```

Next Week

Please bring your laptop.

Week3: Research Designing Social Science
For example, studies in the social sciences focus on methodological reflections on empirical resear ch as well as conceptual integration of chosen methodologies.

- Methodological process
- Research questions
- Source Data: Selection and Procurement
- Preprocessing Preparations
- Analysis