# Chapter 1

Review of Data Science and Econometric

Dr. Woraphon Yamaka

Introduction

- No matter which industry you work in, IT, fashion, food, or finance, there is no doubt that data affects your life and work. At some point in this week

- you will either have or hear a conversation about data.

- In the 19th century, the world was in the grip of the *industrial age*. Mankind was exploring its place in industry alongside giant mechanical inventions

- By the 20th century, we were quite skilled at making huge machines; the goal now was to make them smaller and faster. The industrial age was over and was replaced by what we refer to as the *information age*
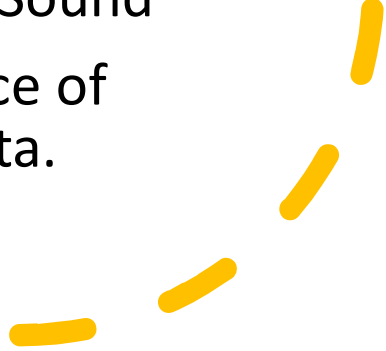
The ENIAC, http://ftp.arl.mil/ftp/historic-computers/
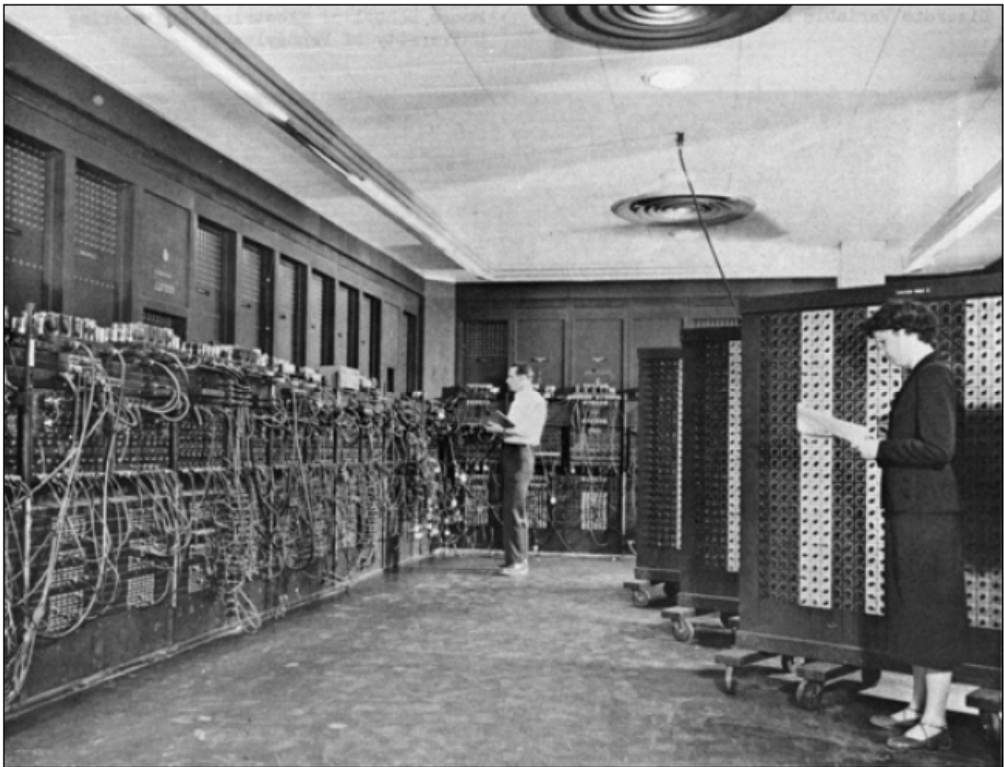
We finally had a d[...]than we could! As with the industrial age, the information age brought us both the good and

Introduction

- We have so much data and we keep making more. We have built insanely tiny machines that collect data 24/7, and it's our job to make sense of it all. Enter the *data age*.
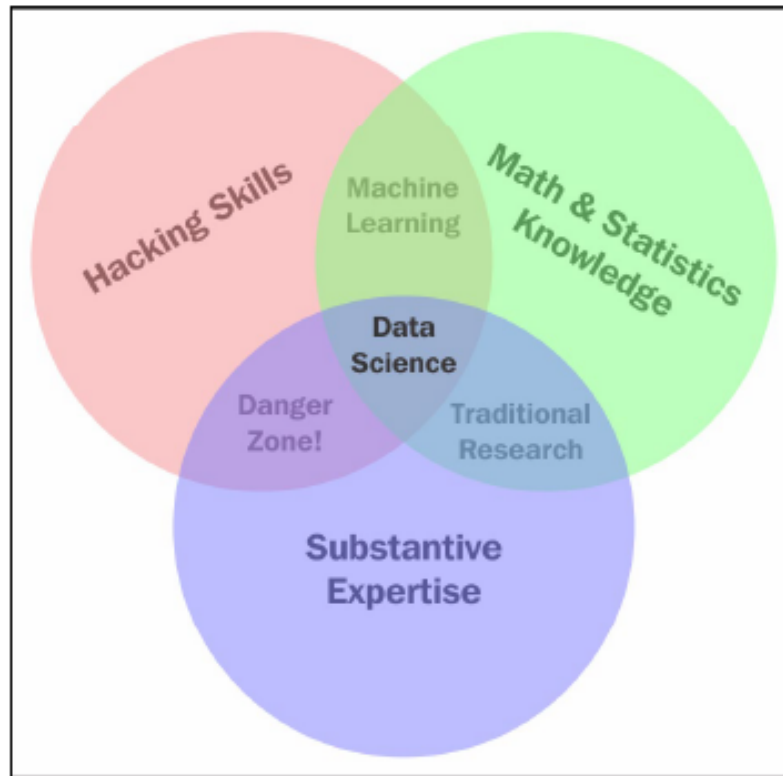
What is data science?

- Let's start by defining what data is. This might seem like a silly first definition to have, but it is very important. Whenever we use the word "data", we refer to a collection of information in either an *organized* or *unorganized* format:

1) **Organized data**: Excel

2) **Unorganized data**:data that is in the free form such as Text, Picture and Sound

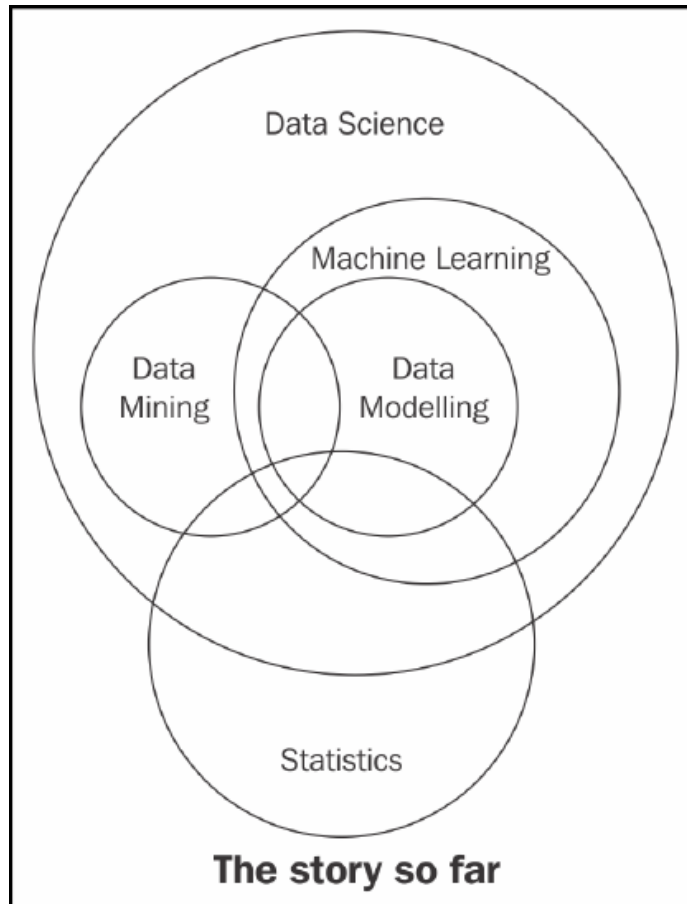- Data science is the art and science of acquiring knowledge through data.

What is the data science do?

- Make decisions
- Predict the future
- Understand the past/present
- Create new industries/products

Understanding data science

The Venn diagram of data science

The story so far

The Five Steps of Data Science

- 1. Asking an interesting question
- 2. Obtaining the data
- 3. Exploring the data
- 4. Modeling the data
- 5. Communicating and visualizing the results

Mathematic, Statistic and Econometrics for Data science

- **Mathematics**
I. Basic symbols/terminology :vector, Graph
II. Logarithms/exponents
III. The set theory : union, intersect
IV. Calculus : differential, Integral, Limit
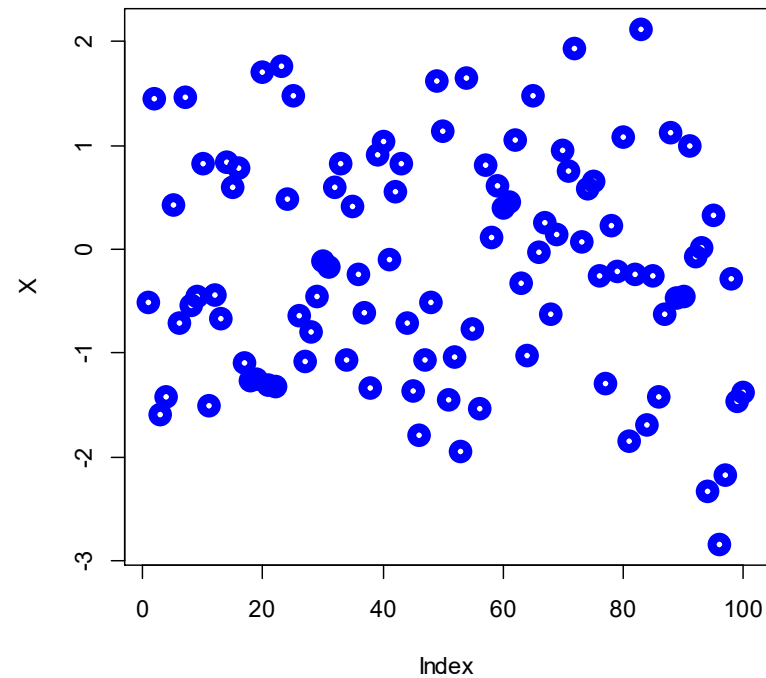V. Matrix (linear) algebra
- **Statistics**
I. Descriptive statistics: mean, max, min, mode, median
II. Inference Statistics : hypothesis testing, sample
- **Econometrics and Machine learning**
I. Modelling : regression, Neural network, Support Vector Machine
II. Testing
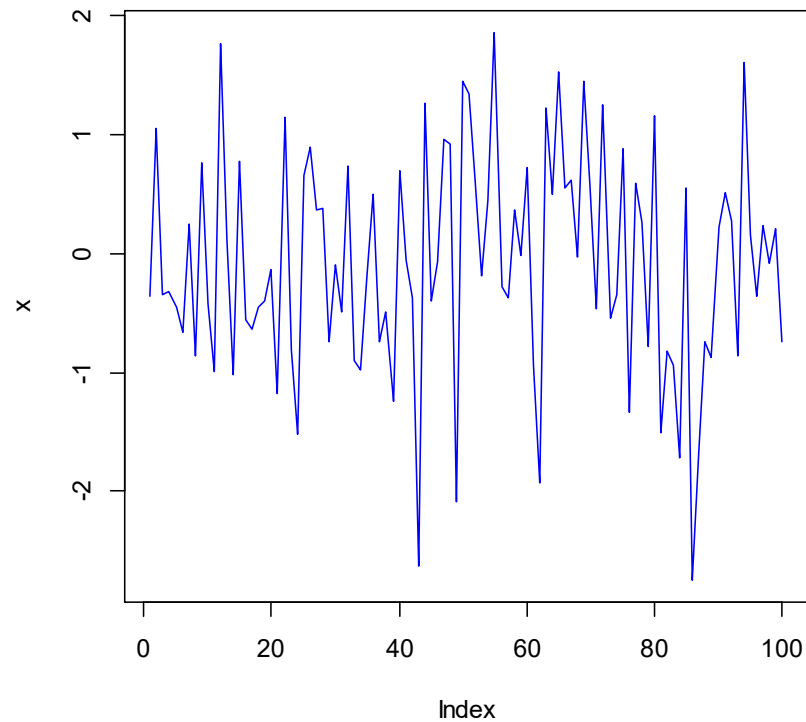
Communicating Data

- **Scatter plots**


```
x=rnorm(100)
plot(x, col="blue", type="p", lwd=10)
```
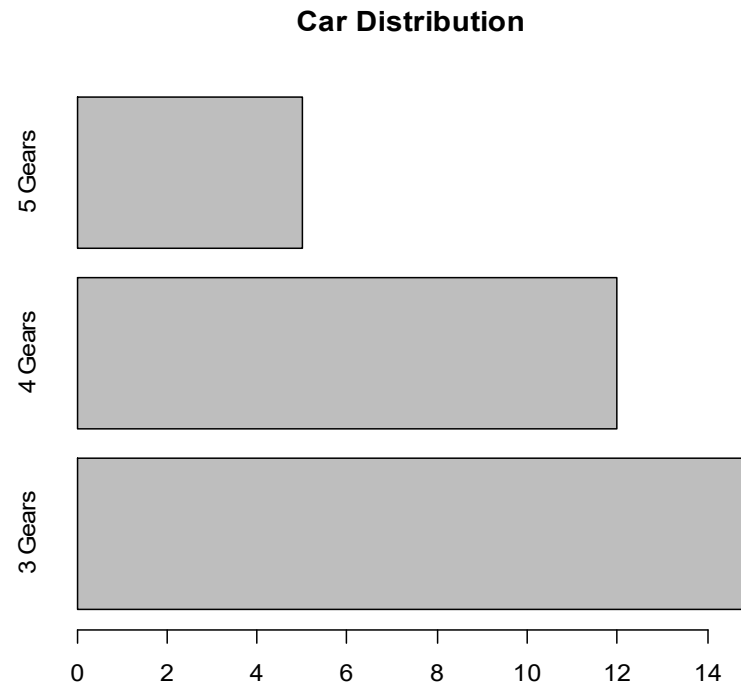
Communicating Data

- **Line graphs**


```
x=rnorm(100)
plot(x, col= "blue", type="l")
```
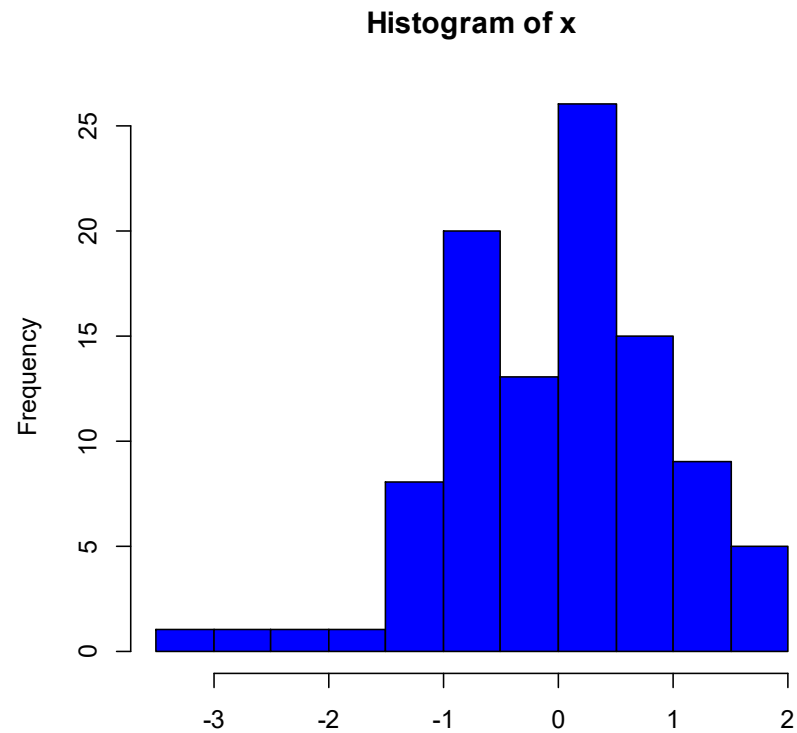
Communicating Data

- **Bar charts**

Car Distribution


```
counts <- table(mtcars$gear)
barplot(counts, main= "Car Distribution", horiz=TRUE, names.arg=c("3 Gears", " "4 Gears",  "5 Gears"))
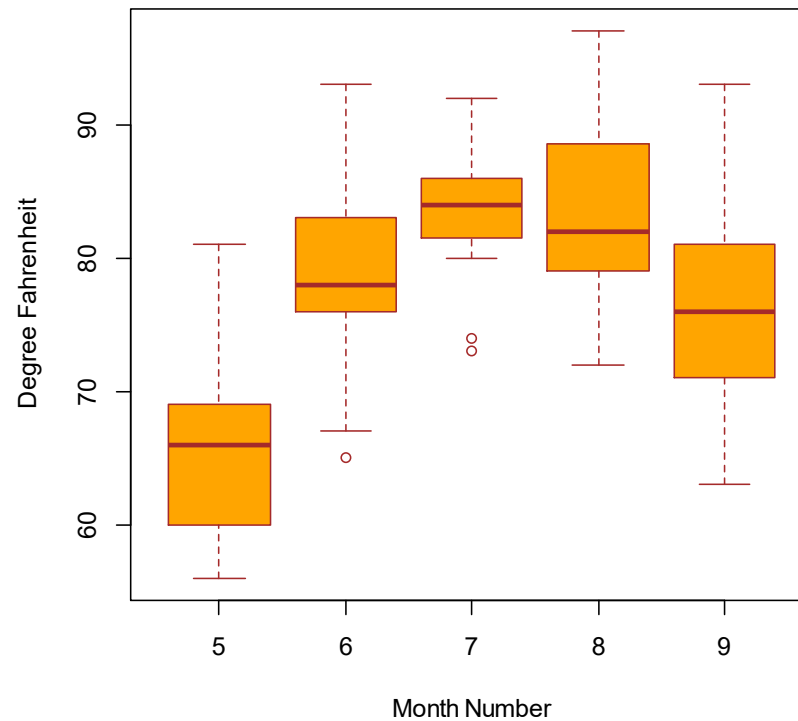```

# Communicating Data

- **Histograms**



Histogram of x

x=rnorm(100)
hist(x, col="blue")

# Communicating Data

- **Box plots**

**Different boxplots for each month**



```
boxplot(Temp ~Month, data=airquality,
main=" Different boxplots for each month", xlab="Month Number" , ylab= "Degree Fahrenheit", col= "orange", border= "brown")
```
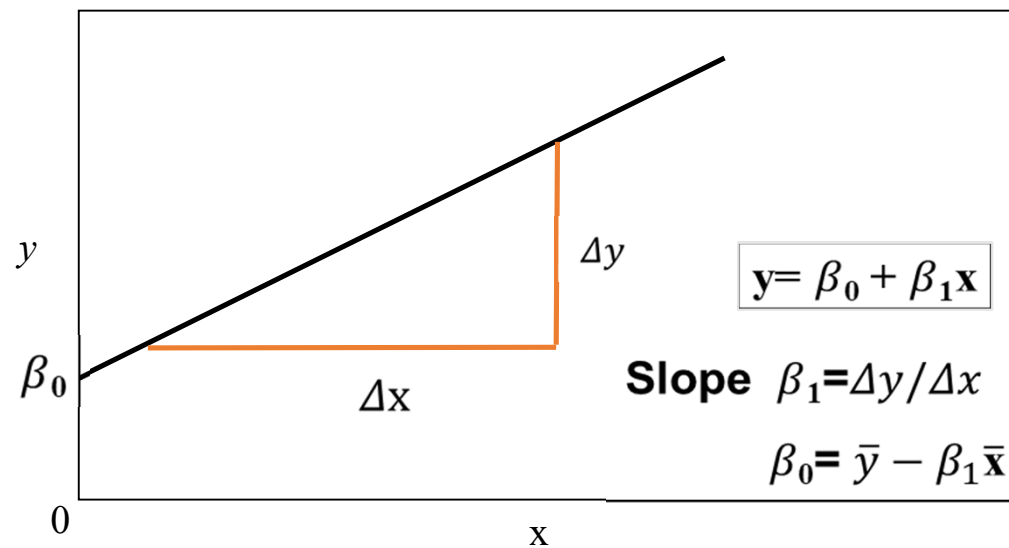
# Other statistics

- Correlation
- Variance
- ANOVA
- Tests

# Econometrics and Machine learning
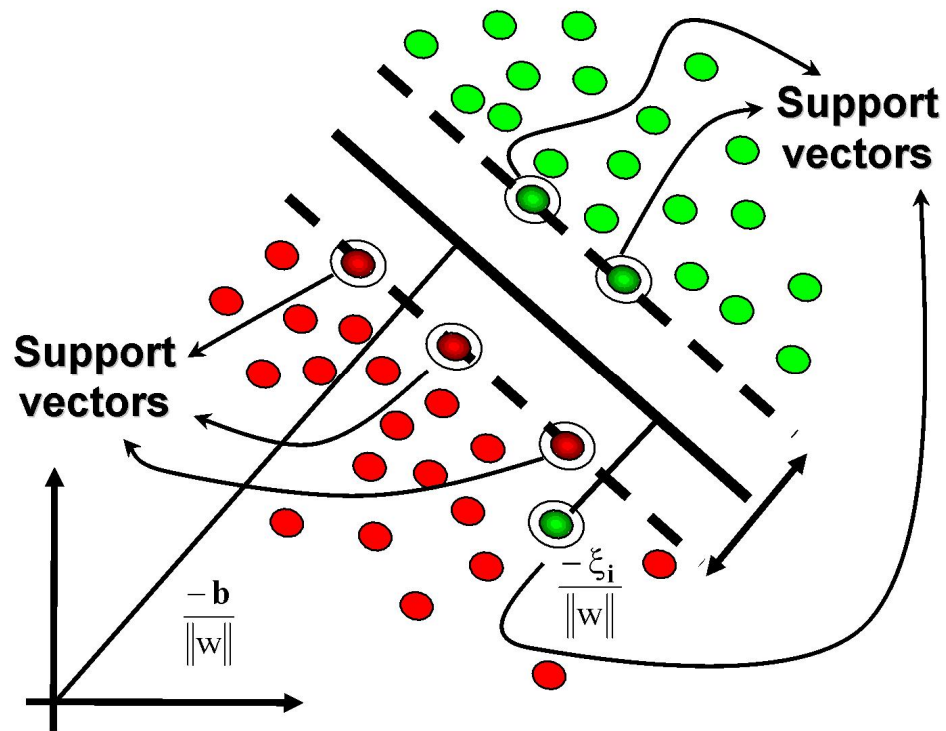
- Linear regression

$$Y = X\beta + u$$



$$y = \beta_0 + \beta_1 \mathbf{x}$$

**Slope** $\beta_1 = \Delta y / \Delta x$

$\beta_0 = \bar{y} - \beta_1 \bar{\mathbf{x}}$

# Econometrics and Machine learning
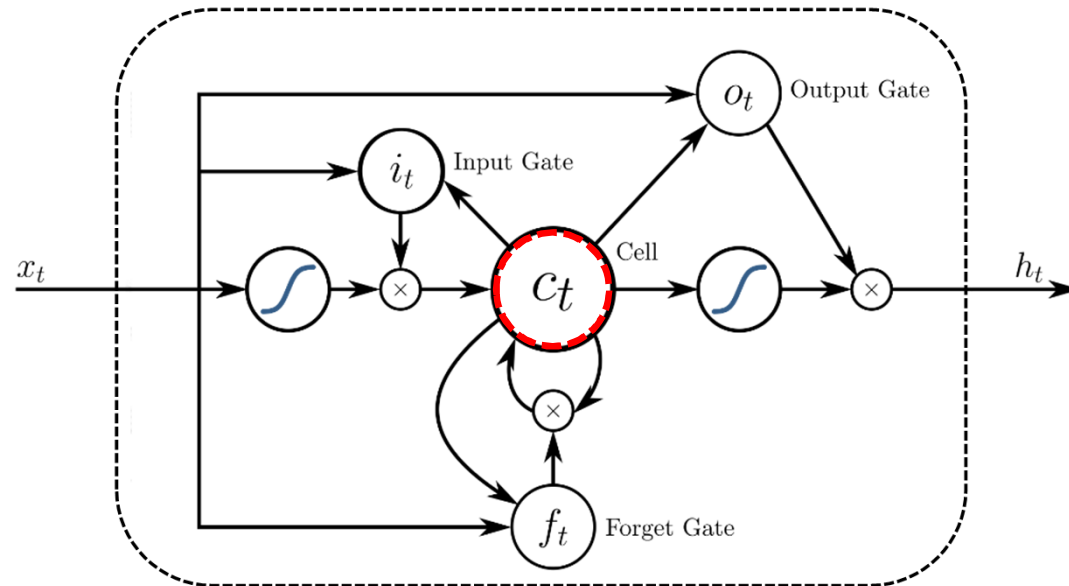
- Support Vector Machine

$$y = \boldsymbol{\omega}'\phi(x_t) + b$$

$y = \boldsymbol{\omega}' \phi(x_t) + b$

# Econometrics and Machine learning

- Neural network

# References

- **Sinan Ozdemir (2016),** Principles of Data Science
- https://www.datamentor.io/r-programming/plot-function/