

# CUHK Taster Fair 2020 BSSc in Data Science and Policy Studies (DSPS)

**Professor Wilson WONG**  
Programme Director of DSPS

**Dr. HO Chi Pui**  
Lecturer of DSPS



The background image shows a laptop screen with a spreadsheet application. The spreadsheet has columns for DATE, DUREE, DISTANCE (MILES), ALLURE (TEMPS/MILE), and REMARQUES. The data rows show dates from 1/4/14 to 21/6/14 with corresponding values. To the right of the spreadsheet is a pie chart with a green and red section, and a table with financial data including 'Dépenses totales' of CHF 2'925. The laptop keyboard is visible in the foreground.

# Content and Overview

## 1. What is DSPS?

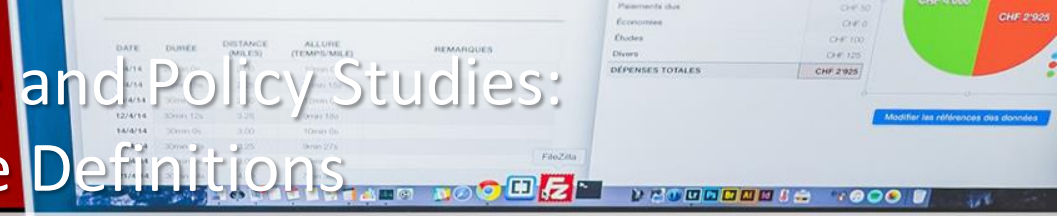
- Scope, Potential Students, Programme Curriculum, Career Prospects

## 2. Case Study

## 3. Admission Requirements

## 4. Q&A

# Data Science and Policy Studies: Some Definitions



- Data Science:
  - an interdisciplinary field that uses scientific and statistical methods, processes, algorithms, and systems, including big data and advanced computer technologies such as AI and IoT (Internet of Things) to generate knowledge and insights from data.
- Policy Studies
  - an interdisciplinary field of study which designs and examines public policy for formulating solutions and strategies to resolve public problems.

DSPS

=

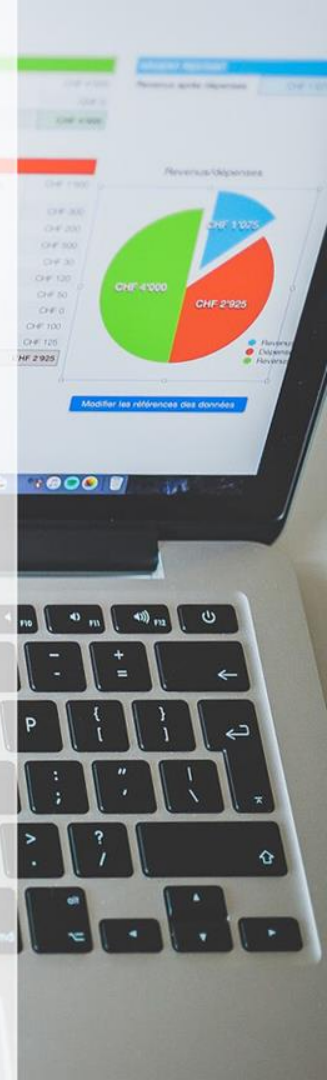
Data Science

×

Social Science & Policy Studies

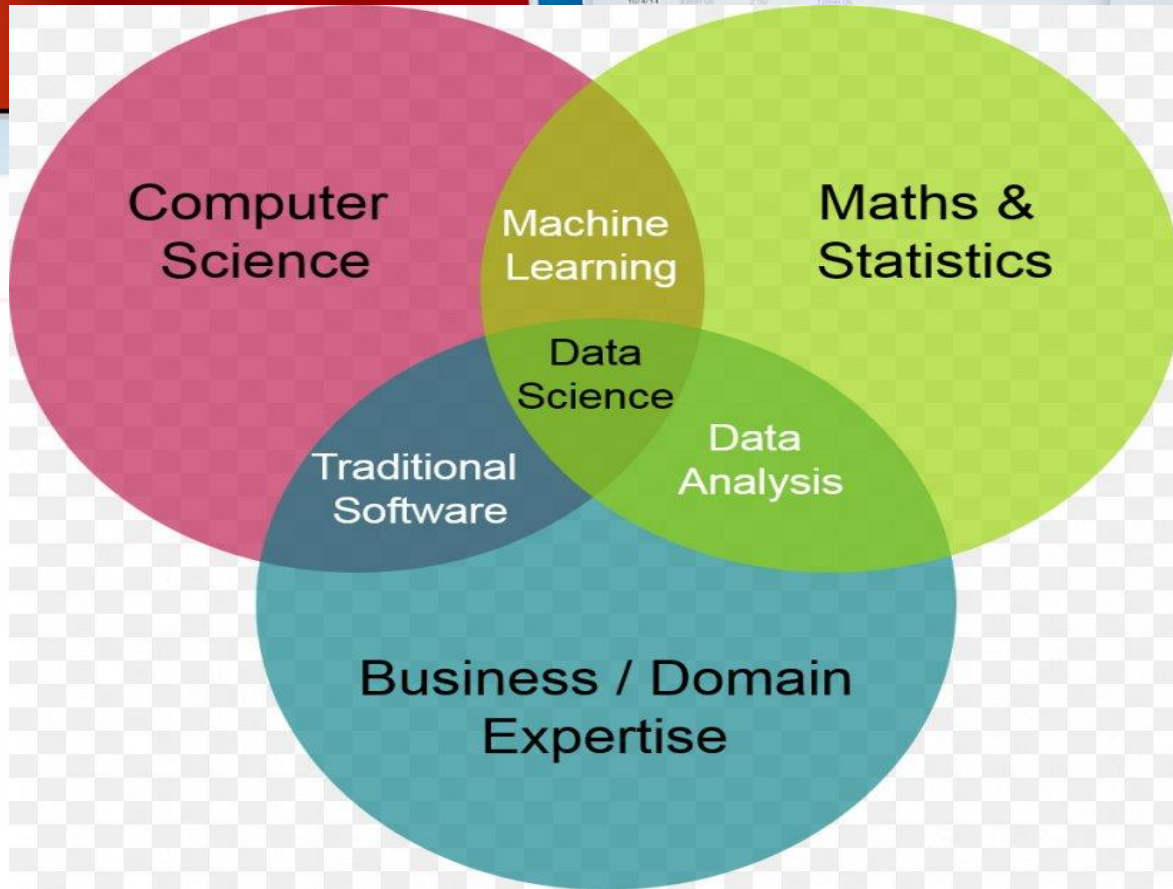
- Knowledge & skills in data science
- Formulate innovative & evidence-based policies

- Ask big questions on our social problems
- Come up with bold answers & initiatives
- Drive impactful changes for social good



# First Question

- Is Data Science only about Data?
  - What is data science
  - What is the process of data science
  - Data science requires subject knowledge for insights and intelligence
    - What data to find?
    - How to interpret them?

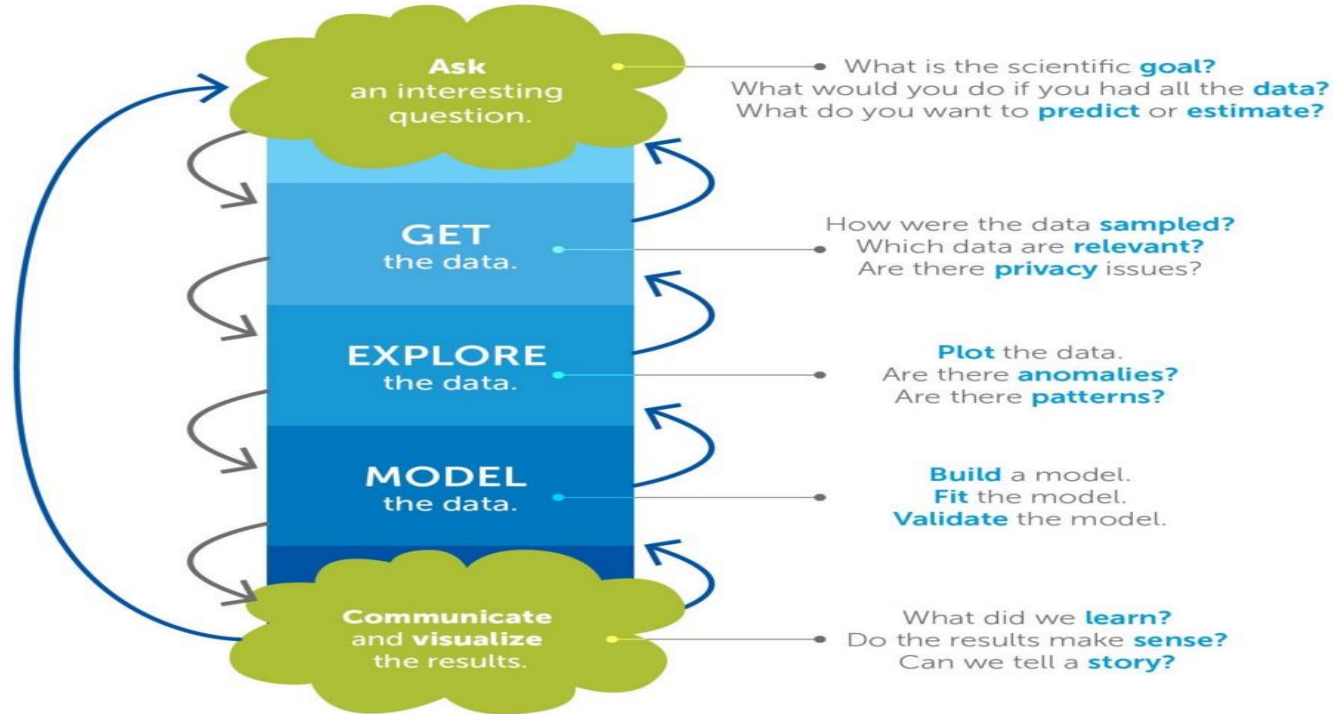


DATE	DUREE	DISTANCE (MILES)	ALLURE (TEMPS/MILE)	REMARQUES
1/4/14	20min 0s	2.00	10min 0s	
3/4/14	25min 10s	2.25	11min 10s	
10/4/14	30min 0s	2.50	12min 0s	

Placements dus	CHF 50
Economies	CHF 0
Échecs	CHF 100
Divers	CHF 125
<b>DÉPENSES TOTALES</b>	<b>CHF 275</b>



# The Data Science Process



Derived from the work of Joe Blitzstein and Hanspeter Pfister, originally created for the Harvard data science course <http://cs109.org/>.

# Potential Students



✓ Are interested in **BOTH** policy studies and data science



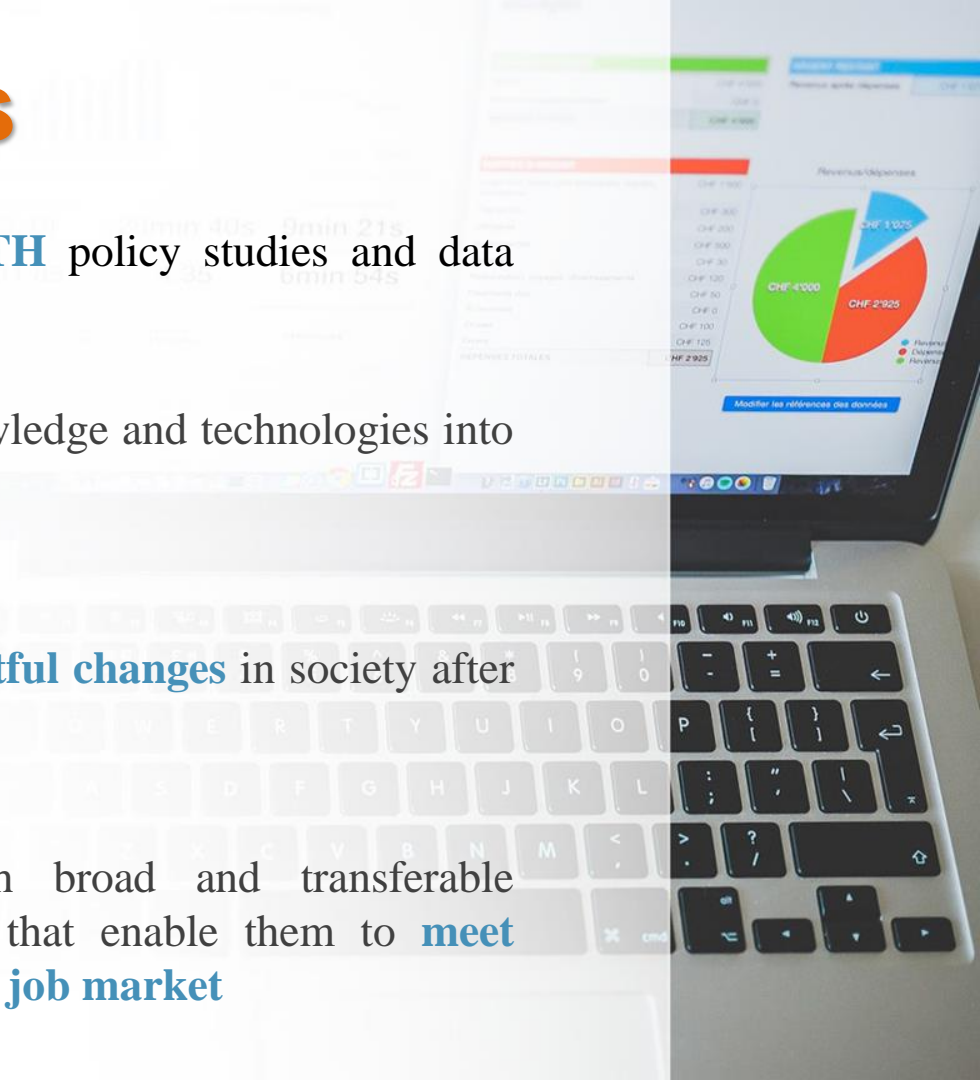
✓ Seek to transform knowledge and technologies into responsible **solutions**



✓ Aspire to **make impactful changes** in society after graduation



✓ Strive to equip with broad and transferable knowledge and skills that enable them to **meet future demands of the job market**





# A Test of Your Interest in Policy Studies

- Are you interested in those issues and current affairs?
- Do you have a policy argument and analysis (evidence-based) to support your position?
  - Government's plan of fiscal package to bail out Ocean Park
    - Support or Not Support?
  - US-China Trade War or New Cold War
    - Who would win? What are the impacts and consequences?
  - Would the national security law weaken Hong Kong's economy and its status as an international financial centre?
    - If so, why? What can be done to avoid it (policy tools and options)?



# Career Prospects

- Students are trained to be policy makers, data analysts, entrepreneurs, consultants and communicators in **public and private sectors**:



✓ Government



- ✓ Think tanks
- ✓ Consulting firms



- ✓ Local and International NGOs



- ✓ Multinational corporations
- ✓ Science & technology companies



# Career Prospects

- Traditional Job Markets + New, Fast-growing and Emerging Markets
- Public Policy and Social Science Training + New Knowledge and Skills of data science and technologies
- The combination is your strength and advantage



# Career Prospects

- Students are supported by faculty members with affiliation / experience in different sectors and organizations, e.g.,
  - Hong Kong SAR Government (Administrative Officer (AO) Grade) and NGOs
  - Central Policy Unit (CPU) / Policy Innovation and Coordination Office (PICO), HKSAR Government
  - Office of the Government Economist
  - Brookings Institution, USA





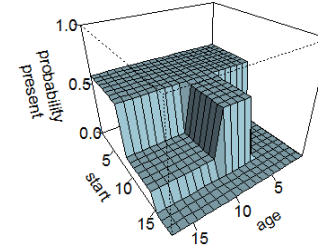
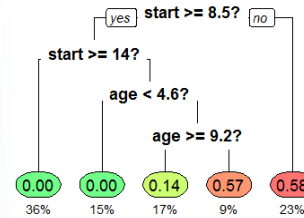
# Technical and analytical skills to be learned in DSPS

## 1. Social science thinking



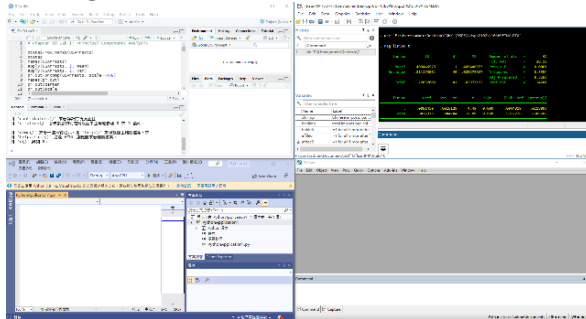
- *Research methods*
- *Big data in social science/policy studies*

## 2. Data science knowledge



- *Statistics, Regression, Statistical learning, Machine learning, Data mining, etc.*

## 3. Usage of statistical software



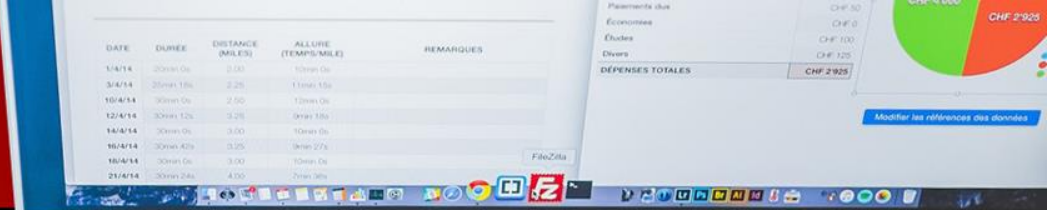
- *Coding in R/Python/STATA*

## 4. Data communication



- *Data/result visualization*





# Taster Fair for DSPTS Programme

Case study:

The U.S. relaxing COVID-19 control measures in late April

# Case study: COVID-19 in the U.S.

JOHNS HOPKINS  
UNIVERSITY OF MEDICINE

CORONAVIRUS  
RESOURCE CENTER

Home Maps & Trends Testing Tracing News & Information COVID-19 Basics Videos & Live Events

World Map

U.S. Map

Critical Trends

COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)

Total Confirmed

2,935,008

Confirmed Cases by  
Country/Region/Sovereignty

2,935,008 US
1,623,284 Brazil
697,413 India
686,777 Russia
305,703 Peru
298,557 Chile
287,290 United Kingdom
261,750 Mexico
251,789 Spain
243,051 Iran
241,819 Italy
231,818 Pakistan
213,716 Saudi Arabia
206,844 Turkey
205,721 South Africa



Global Deaths

130,277

130,277 deaths  
US

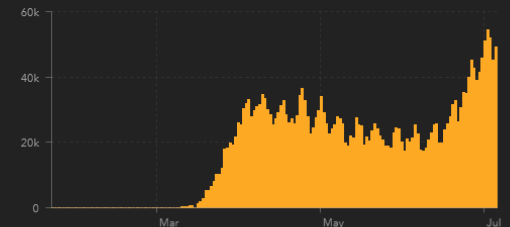
US State Level  
Deaths, Recovered

32,219 deaths, 71,040 recovered New York US
15,229 deaths, 30,729 recovered New Jersey US
8,198 deaths, 93,157 recovered Massachusetts US
7,026 deaths, recovered Illinois US
6,754 deaths, 70,437 recovered Pennsylvania US
6,440 deaths, recovered California US
6,221 deaths, 52,841 recovered Michigan US

Global Deaths

Global Recovered

US Deaths, Recovered



188

countries/regions

Lancet Inf Dis Article: [Here](#). Mobile Version: [Here](#).

Lead by JHU CSSE. Technical Support: [Esri Living Atlas team](#) and [JHU APL](#). Financial Support: [JHU](#) and [NSF](#). Resource support: [Slack](#), [Github](#) and [AWS](#). Click [here](#) to donate to the CSSE dashboard team, and other JHU COVID-19 Research Efforts. [FAQ](#). Read more in [this blog](#). [Contact Us](#).

Last Updated at (M/D/YYYY)

7/7/2020, 8:33:59 AM

# Case study: COVID-19 in the U.S.

The U.S. began to relax its COVID-19 control measures in late April.  
Was this sound?

rthk.hk 香港電台網站 中文新聞

新聞主頁 即時新聞 視像新聞 新聞專題 新聞節目 新聞圖片 新聞簡報

2020.07.06 星期一 + 30°C • 78%

即時新聞 主頁 > 即時新聞 > 國際

## 美國民調：近六成人認為目前放寬限制措施過早

2020-04-20 HKT 21:10

分享工具

美國至今確診新冠肺炎人數突破76萬人，單日再新增超過2萬宗確診，累計超過4萬人死亡。自白宮宣布重啟經濟措施後，周日華盛頓、加州等地均有民眾上街，繼續要求當局解除「居家令」。

當地有最新民調顯示，58%受訪者認為目前放寬限制措施是太早，32%人則認為來得太慢；另外，大約7成受訪者憂慮家人受感染，較一個月前的同類調查，增加20個百分點。

## 美國各州放寬限制 加快重啟經濟腳步

分享 留言 列印 存新聞

2020-04-27 14:22 聯合晚報 記者薛伶如、編譯樊基/綜合報導



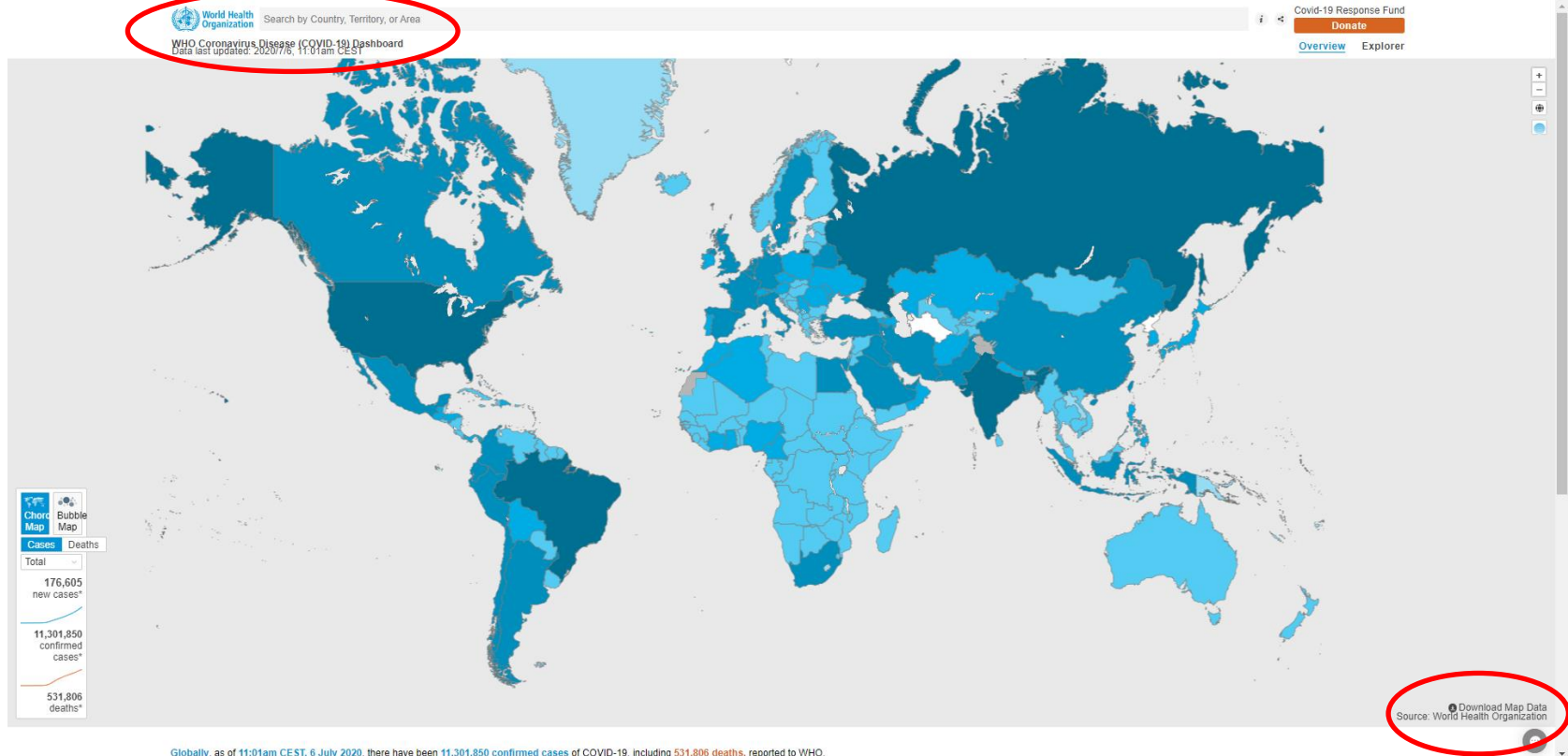
美國總統川普已經減少出席白宮疫情簡報的時間，因為他在23日的記者會上發言。 歐新社

Sound policy?



# Case study: COVID-19 in the U.S.

- Step 1: Data collection



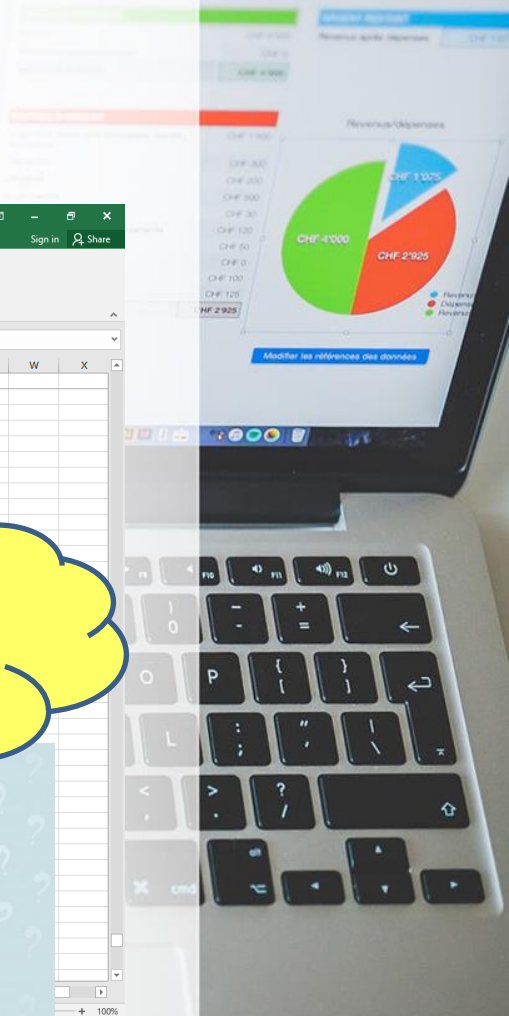
Globally, as of 11:01am CEST, 6 July 2020, there have been 11,301,850 confirmed cases of COVID-19, including 531,806 deaths, reported to WHO.

# Case study: COVID-19 in the U.S.

- Step 1: Social science thinking

Date_report	Country	Country	WHO region	New_cases	Cumulative_cases	New_deaths	Cumulative_deaths
25/3/2020	US	United States of America	AMRO	18093	103321	425	1668
25/7/2020	US	United States of America	AMRO	19332	122653	444	2112
25/7/2020	US	United States of America	AMRO	17987	140640	286	2398
25/7/2020	US	United States of America	AMRO	22559	163199	452	2850
25/7/2020	US	United States of America	AMRO	24103	187302	996	3846
25/7/2020	US	United States of America	AMRO	26298	213600	947	4793
25/7/2020	US	United States of America	AMRO	28103	241703	1061	5854
25/7/2020	US	United States of America	AMRO	32105	273808	1166	7020
25/7/2020	US	United States of America	AMRO	33510	307318	1338	8358
25/7/2020	US	United States of America	AMRO	26493	333811	1201	9559
25/7/2020	US	United States of America	AMRO	29510	363321	1286	10845
25/7/2020	US	United States of America	AMRO	31709	395030	1895	12740
25/7/2020	US	United States of America	AMRO	30859	425889	1925	14665
25/7/2020	US	United States of America	AMRO	35386	461275	1931	16596
25/7/2020	US	United States of America	AMRO	31606	492881	1920	18516
25/7/2020	US	United States of America	AMRO	31633	524514	1928	20444
25/7/2020	US	United States of America	AMRO	29308	553822	1528	21972
25/7/2020	US	United States of America	AMRO	24446	578268	1504	23476
25/7/2020	US	United States of America	AMRO	25802	604070	2395	25871
25/7/2020	US	United States of America	AMRO	28711	632781	6409	32280
25/7/2020	US	United States of America	AMRO	32549	665330	2018	34298
25/7/2020	US	United States of America	AMRO	30023	695353	2438	36736
25/7/2020	US	United States of America	AMRO	28252	723605	1776	38512
25/7/2020	US	United States of America	AMRO	27668	751273	1801	40313
25/7/2020	US	United States of America	AMRO	25634	776907	2154	42467
25/7/2020	US	United States of America	AMRO	24019	800926	2575	45042
25/7/2020	US	United States of America	AMRO	29127	830053	2390	47432
25/7/2020	US	United States of America	AMRO	30719	860772	1742	49174
25/7/2020	US	United States of America	AMRO	38509	899281	2151	51325
25/7/2020	US	United States of America	AMRO	32417	931698	1964	53289
25/7/2020	US	United States of America	AMRO	29218	960916	1109	54398
25/7/2020	US	United States of America	AMRO	22541	983457	1322	55720
25/7/2020	US	United States of America	AMRO	20517	1003974	2010	57730
25/7/2020	US	United States of America	AMRO	31379	1035353	2902	60632
25/7/2020	US	United States of America	AMRO	31774	1067127	2005	62637
25/7/2020	US	United States of America	AMRO	2675	1093880	5000	67637
25/7/2020	US	United States of America	AMRO	31119	1125719	156	67793
25/7/2020	US	United States of America	AMRO	29266	1154985	-514	67279

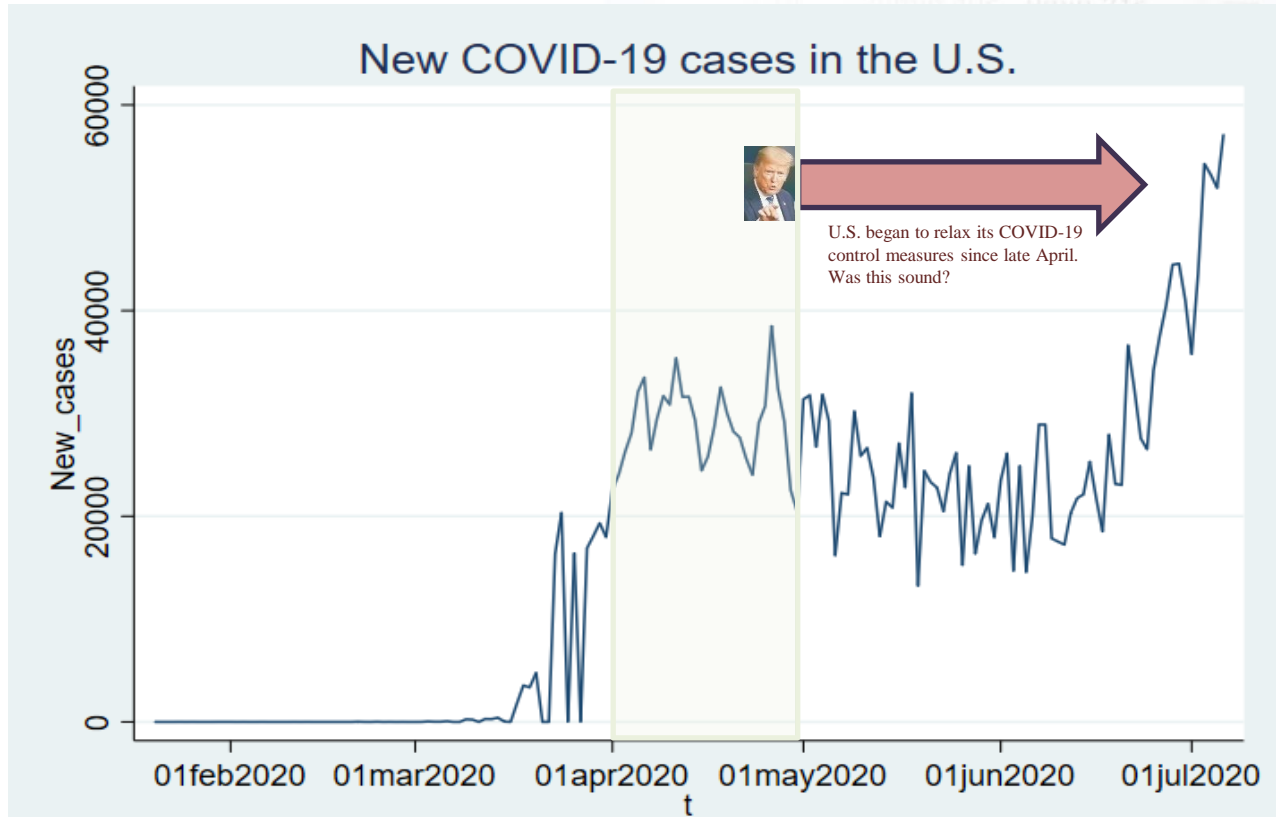
The U.S. should relax the COVID-19 measures only when there was strong evidence of declining new cases...





# Case study: COVID-19 in the U.S.

- Step 1: Research question



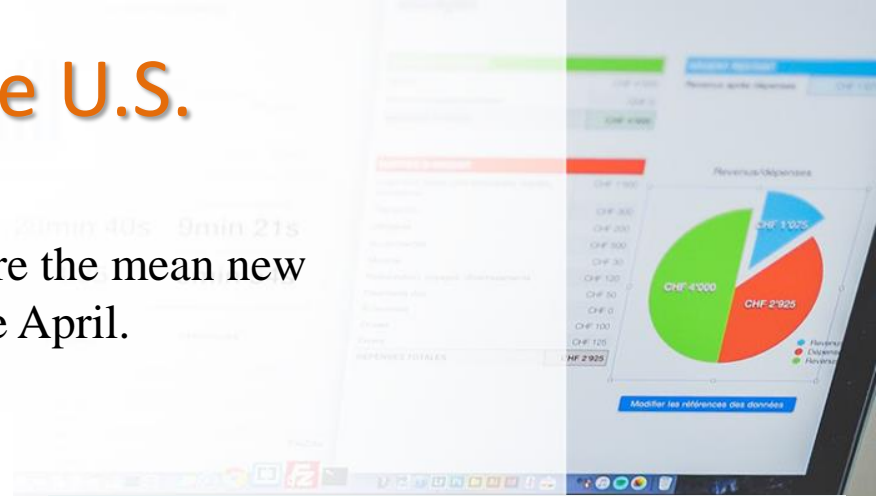
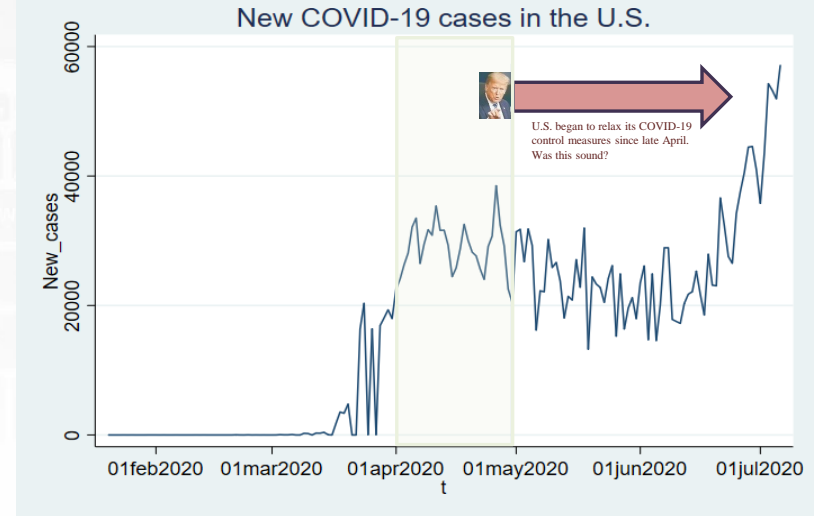
# Case study: COVID-19 in the U.S.

- Step 2: Data science knowledge
- We can employ *hypothesis testing* to compare the mean new COVID-19 cases in the U.S. in early and late April.

## Steps for Hypothesis Testing

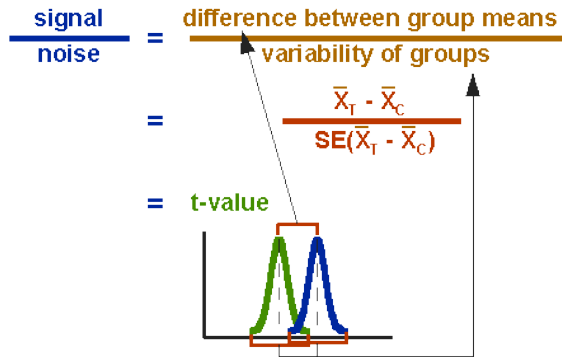
### HYPOTHESIS TESTING PROCEDURE

1. Start with a well-developed, clear research problem or question
2. Establish hypotheses, both null and alternative
3. Determine appropriate statistical test and sampling distribution
4. Choose the Type I error rate
5. State the decision rule
6. Gather sample data
7. Calculate test statistics
8. State statistical conclusion
9. Make decision or inference based on conclusion



# Case study: COVID-19 in the U.S.

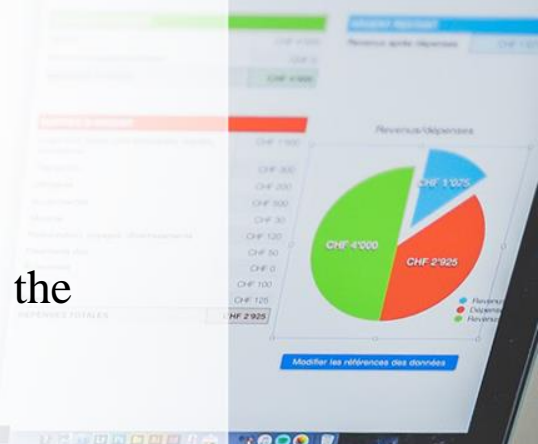
- Step 2: Data science knowledge
- In the process we will apply statistical formula – in this case the *t-statistics* to test difference in mean between two samples.



$$t = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Where  $\bar{x} = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i, \bar{y} = \frac{1}{n_2} \sum_{j=1}^{n_2} y_j$

$$S^2 = \frac{1}{n_1 + n_2 - 2} \left[ \sum_i (x_i - \bar{x})^2 + \sum_j (y_j - \bar{y})^2 \right]$$



# Case study: COVID-19 in the U.S.

- Step 3: Usage of statistical software

The image displays a Stata 16.0 interface with a Do-file Editor and a Graph window. The Do-file Editor contains the following code:

```
do "F:\Users\admin\Desktop\US_covid19.do"
use "F:\Users\admin\Desktop\US_covid19.dta", clear
gen t=td(20jan2020)+_n-1
format t %td
tsset t
time variable: t, 20jan2020 to 06jul2020
delta: 1 day

// test whether covid-19 situation improves in late Apr relative to early Apr
gen new_case_earlyapr = New_case if inrange(t, td(1apr2020), td(15apr2020))
(154 missing values generated)

gen new_case_lateapr = New_case if inrange(t, td(16apr2020), td(30apr2020))
(154 missing values generated)

ttest new_case_earlyapr = new_case_lateapr, unpaired

Two-sample t test with equal variances
```

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]
new_yapr	15	29175.2	965.5288	3739.477	27104.35 31246.05
new_eapr	15	28380.4	1145.309	4435.761	25923.96 30836.84
combined	30	28777.8	739.6591	4851.28	27265.03 30290.57
diff		794.8	1497.991		-2273.696 3863.296

```
diff = mean(new_case_earlyr) - mean(new_case_lateapr)      t = 0.5306
Ho: diff = 0                      degrees of freedom = 28

Ha: diff < 0                      Ha: diff != 0              Ha: diff > 0
Pr(T < t) = 0.7001                  Pr(|T| > |t|) = 0.5999                  Pr(T > t) = 0.2999

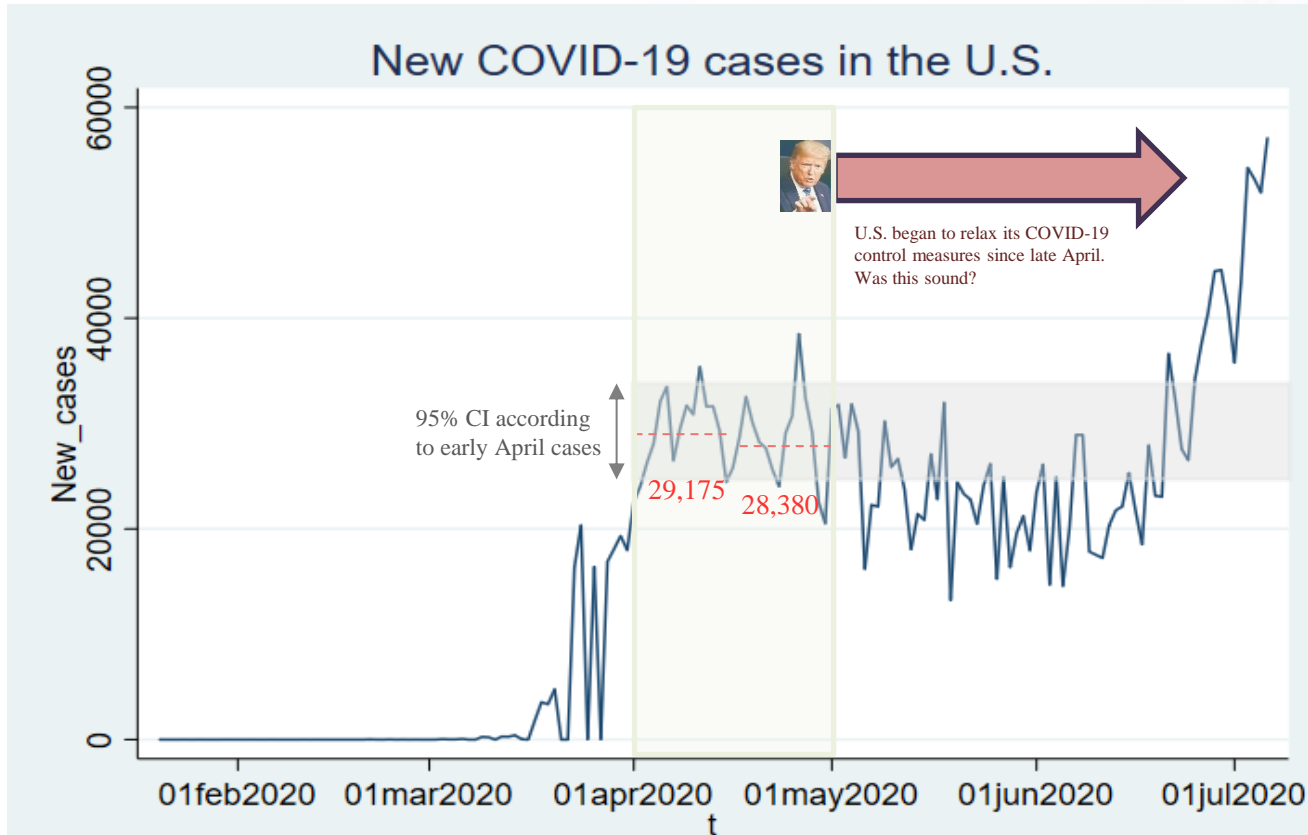
tsline New_cases, title(New COVID-19 cases in the U.S.)

end of do-file
```

The Graph window displays a line plot titled "New COVID-19 cases in the U.S." showing the number of new cases over time from February 1, 2020, to July 1, 2020. The y-axis is labeled "New\_Cases" and ranges from 0 to 60,000. The x-axis is labeled "t" and shows dates from 01feb2020 to 01jul2020. The plot shows a sharp increase in cases starting in late April, peaking in late May, and then fluctuating with an overall upward trend through July.

# Case study: COVID-19 in the U.S.

- Step 4: Data/Result Visualization



I see!





# Case study: COVID-19 in the U.S.

- But were there other considerations on whether to relieve the COVID-19 control measures in late-Apr?
  - Economic activities?
  - Data for economic performance? GDP, Industrial production

## U.S. Economy Sees Sharp Downturn Amid COVID-19 Crisis

Quarterly real GDP growth in the United States\*



\* percent change from preceding quarter; seasonally adjusted at annual rates

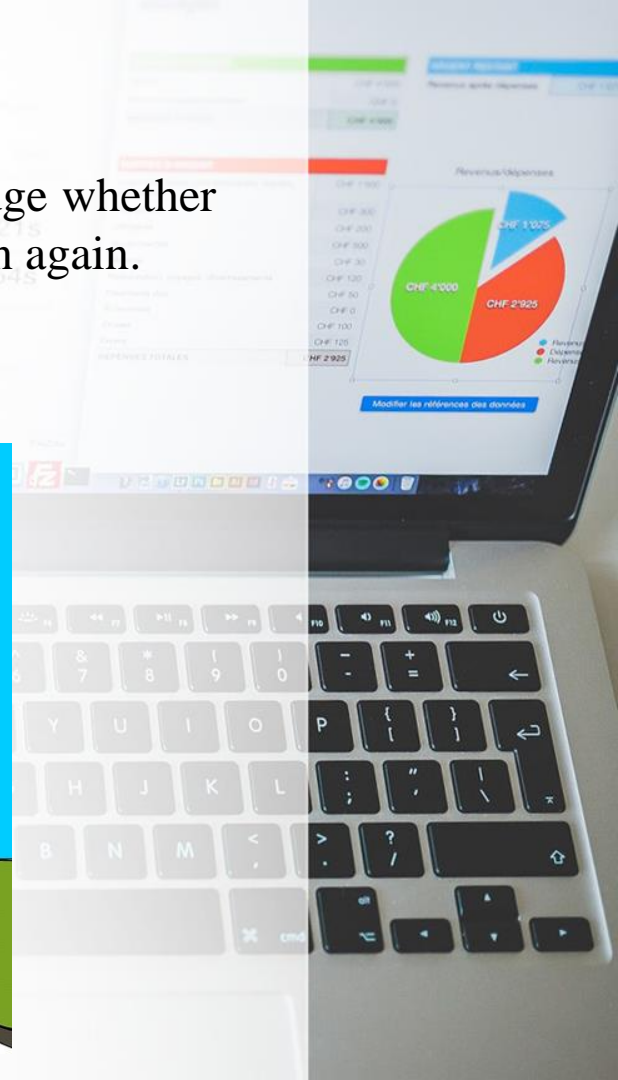
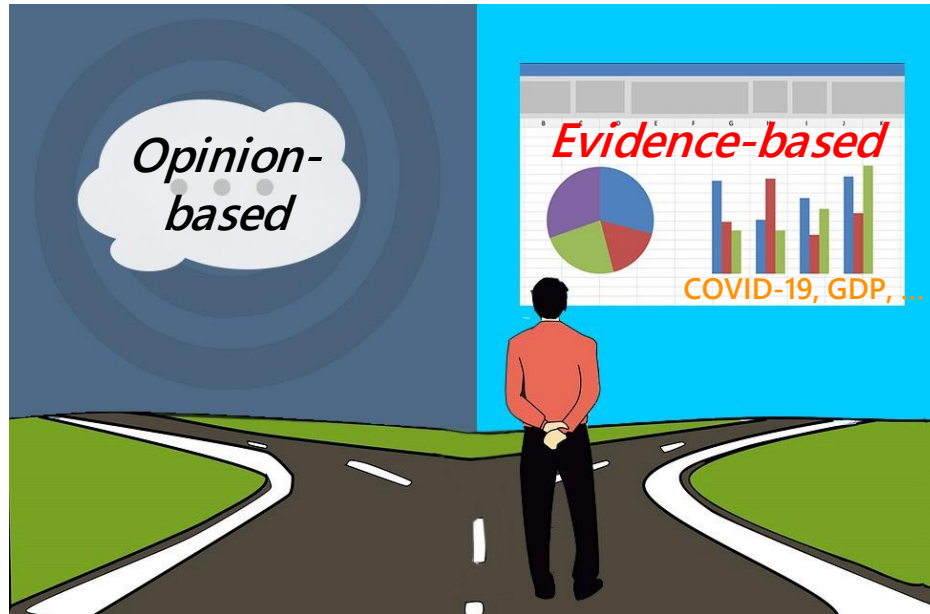
Source: U.S. Bureau of Economic Analysis



statista

# Case study: COVID-19 in the U.S.

- How to weigh public health and economic security to judge whether the policy decision was sound? Go back to research design again.
  - Social science thinking
  - Policy studies knowledge



# Admissions Requirements - JUPAS

## Minimum University Requirements for JUPAS Applicants

They shall have obtained in the HKDSE Examination:

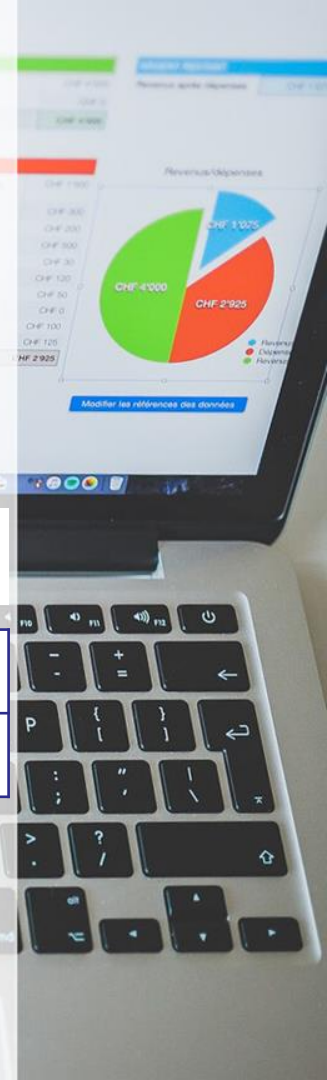
### Calculation of Admission Scores

**Best 5 Scores + Bonus Points** (6<sup>th</sup> & 7<sup>th</sup> Subject)

<u>Subject</u>	<u>Level</u>
Chinese Language	<b>3</b>
English Language	<b>3</b>
Mathematics	<b>3</b>
Liberal Studies	<b>2</b>
2 Elective Subjects <u>OR</u> 1 Elective Subject plus Mathematics Extended Part Module 1 or 2	<b>3</b>

### Grade Point Conversion for Category A: Core & Elective Subjects

<u>Level</u>	5**	5*	5	4	3	2	1
<u>Score</u>	8.5	7	5.5	4	3	2	1



# Admissions Requirements - JUPAS

## 2019 DSPS Admissions Scores (for Reference only)

<u>Subject</u>	<u>Upper Quartile</u>	<u>Median</u>	<u>Lower Quartile</u>
Total Reference Score^	26	24	24

The Total Reference Score is the total score of the **Best 5 Subject**  
(where level 5\*\* = 7, level 5\* = 6, level 5 = 5, level 4 = 4, level 3 = 3, level 2 = 2, level 1 = 1)



# Programme Curriculum



	Units
Major Required Courses (e.g. Foundation of Data Science, Data Science and Public Policy)	27
Resident Study Overseas	9
Internship	3
Capstone Courses	6
Major Elective Courses (e.g. Social Science Disciplinary Courses OR Selected Technology & Statistics Courses offered by Faculty of Science and Faculty of Engineering)	18
Faculty Package	9
<b>Total</b>	<b>72</b>



# Course Pattern (Recommended)



First Year	<p>DSPS Required Courses:</p> <ul style="list-style-type: none"> <li>• DSPS1001 Introduction to Policy Sciences</li> <li>• DSPS1002 Data Science and Public Policy</li> <li>• DSPS1003 Foundation of Data Science</li> </ul> <p>Faculty Package</p>
Second Year	<p>DSPS Required Courses:</p> <ul style="list-style-type: none"> <li>• DSPS2101 Research Methods for Policy Studies</li> <li>• DSPS2102 Statistical Analysis for Policy Decision</li> <li>• DSPS2201 Data Analytics for Public Policy I</li> <li>• DSPS2301 Policy Analysis and Design Thinking</li> <li>• DSPS2501 Managing Technology and Policy Innovation</li> </ul> <p>Major Elective Courses</p>
Summer of Second Year	<ul style="list-style-type: none"> <li>• <b>DSPS3801 Internship</b></li> </ul>

Third Year	<p><b>Compulsory Resident Study Overseas (Term 1)</b></p> <p>DSPS Required Courses:</p> <ul style="list-style-type: none"> <li>• DSPS3202 Data Analytics for Public Policy II</li> <li>• DSPS3501 Policy Leadership and Entrepreneurship Workshop</li> </ul> <p>Major Elective Courses</p>
Fourth Year	<p>DSPS Required Courses:</p> <ul style="list-style-type: none"> <li>• <b>DSPS4801 Graduation Capstone Project I</b></li> <li>• <b>DSPS4802 Graduation Capstone Project II</b></li> </ul> <p>Major Elective Courses</p>

DATE	DUREE	DISTANCE (MILES)	ALLURE (TEMPS/MILE)	REMARQUES
1/4/14	20min 0s	2.00	10min 0s	
3/4/14	25min 18s	2.25	11min 15s	
10/4/14	30min 0s	2.50	12min 0s	
12/4/14	30min 12s	3.25	09min 58s	
14/4/14	30min 0s	3.00	10min 0s	
16/4/14	30min 42s	3.25	09min 27s	
16/4/14	30min 0s	3.00	10min 0s	
21/6/14	30min 36s	4.00	07min 36s	

FinZita

Placements dus CHF 50  
Economies CHF 0  
Echecs CHF 100  
Divers CHF 125  
DEPENSES TOTALES CHF 2,925

CHF 9,500  
CHF 2,925

Modifier les références des données

# Q&A



DATE	DUREE	DISTANCE (MILES)	ALLURE (TEMPS/MILE)	REMARQUES
1/4/14	20min Co.	2.00	10min Co.	
3/4/14	20min 10s	2.25	10min 10s	
10/4/14	20min Co.	2.50	10min Co.	
12/4/14	20min 10s	2.75	10min 10s	

Placements dus CHF: 50  
Economies CHF: 0  
Dépenses CHF: 100  
Dépenses CHF: 125  
DÉPENSES TOTALES CHF 2925

CHF 4 000  
CHF 2 925

Data Science and Policy Studies (DSPS)  
Room 516, Chen Kou Bun Building,  
Chung Chi Campus, CUHK  
Tel: (852) 3943 4757  
Email: [dspd@cuhk.edu.hk](mailto:dspd@cuhk.edu.hk)  
Website: [dspd.ssc.cuhk.edu.hk](http://dspd.ssc.cuhk.edu.hk)

 @dspscuhk  @dspscuhk



# Thank you!